华章数学译丛

15

Introduction to Linear Programming

线性规划导论

（美）Leonid Nison Vaserstein
Christopher Cattelier Byrne 著

谢金星　姜启源　张立平　等译



机械工业出版社
China Machine Press

HZ BOOKS

PEARSON Prentice Hall

线性规划导论

这本教材是从一门高级的本科生数学课程演变而来的。本书覆盖了线性规划的基本主题，同时包含许多高级主题。通过不同难易程度的习题，为学生提供逐步提高的机会，并使优秀的学生完成更具挑战性的任务。

本 书 特 点

- 强调逻辑和应用建模，使得管理、经济和运筹学等非数学专业的学生能在各自的领域中学习使用线性规划。
- 应用数学工具，但不依赖严密的数学研究进展，为数学专业的高年级学生提供进一步研究所需的理论。
- 难易不同的习题为不同程度的学生提供具有挑战性的任务。
- 附录介绍有关数学规划更深入的材料，为数学专业学生提供高级主题，为面向应用的学生提供已知方法的参考。
- 论述建模问题的系统方法论，为学生指明如何将建模发展为一种技能，而不是依赖他们的直觉解决问题。
- 利用计算机软件解决线性规划问题，指导学生利用各种计算机软件包以及在线软件解决线性规划问题。

本书的习题答案可从**http://www.math.psu.edu/vstein/LPbook/sol2.pdf**上下载，也可登录华章网站。

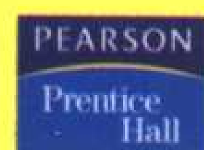
作 者 简 介

Leonid Nison Vaserstein 于1966年在莫斯科国立大学获得数学硕士学位，1969年在该校获得数学博士学位。1984~1985年是普林斯顿高级研究所的成员，曾担任几种数学期刊和文摘杂志的评论员。现任宾夕法尼亚州立大学数学教授。



Christopher Cattelier Byrne 宾夕法尼亚州立大学数学教授。

Introduction to Linear Programming



www.PearsonEd.com



影印版
ISBN 7-111-15893-3
定价：39.00 元



封面设计 · 杨宇梅



华章图书

华章网站 http://www.hzbook.com

网上购书：www.china-pub.com

投稿热线：(010) 88379604
购书热线：(010) 68995259, 68995264
读者信箱：hzjsj@hzbook.com

ISBN 7-111-17329-5/O · 455
定价：33.00 元





华 章 数 学 译 丛

15

Introduction to Linear Programming

线性规划导论

(美) Leonid Nison Vaserstein
Christopher Cattelier Byrne 著

谢金星 姜启源 张立平 等译



机械工业出版社
China Machine Press

线性规划是运筹学的一个分支，它是最优化问题领域中最简单、最基本和使用最广泛的方法，广泛应用于工农业、军事、交通运输、决策管理与规划、科学实验等领域.

本书是从一门高级的本科生数学课程演变而来的，涵盖了线性规划的基本主题，同时包含许多高级主题. 此外，书中给出了很多例子以及难易程度不同的习题，为学生提供逐步提高的机会.

本书从入门开始介绍，只假定读者具有很少的数学基础，因此适用于具有不同数学基础和来自不同专业的学生，包括数学、计算机科学、统计学、工程科学、精算学、计算机工程、工商学管理专业的学生.

Simplified Chinese edition copyright © 2005 by Pearson Education Asia Limited and China Machine Press.

Original English language title: *Introduction to Linear Programming* (ISBN 0-13-035917-3) by Leonid Nison Vaserstein and Christopher Cattelier Byrne, Copyright © 2003.

All rights reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Prentice Hall.

本书封面贴有 Pearson Education(培生教育出版集团)激光防伪标签，无标签者不得销售.

版权所有，侵权必究.

本书法律顾问　北京市展达律师事务所

本书版权登记号：图字：01-2005-0527

图书在版编目(CIP)数据

线性规划导论/(美)瓦泽斯坦(Vaserstein, L. N.)等著；谢金星等译. —北京：机械工业出版社，2006.1

(华章数学译丛)

书名原文：Introduction to Linear Programming

ISBN 7-111-17329-5

Ⅰ. 线…　Ⅱ. ①瓦…　②谢…　Ⅲ. 线性规划　Ⅳ. O221.1

中国版本图书馆 CIP 数据核字(2005)第 102230 号

机械工业出版社(北京市西城区百万庄大街 22 号　邮政编码　100037)

责任编辑：王敏娟　迟振春

北京诚信伟业印刷有限公司印刷·新华书店北京发行所发行

2006 年 1 月第 1 版第 1 次印刷

787mm×1020mm 1/16 · 16.5 印张

印数：0 001-4 000 册

定价：33.00 元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线：(010)68326294



华章教育　赤诚奉献

经典原版书库（相关部分）

具体数学：计算机科学基础（第2版） Graham/49.00
组合数学（第3版） Brualdi/35.00
离散数学及其应用（第5版） Rosen/79.00
离散数学及其应用（第4版） Rosen/59.00
高等微积分 Fitzpatrick/69.00
金融数学 Stampfli/35.00
数值分析（第3版） Kincaid/75.00
概率论及其在投资、保险、工程中的应用 Bean/55.00
代数 Isaacs/65.00
傅里叶分析与小波分析导论 Pinsky/49.00
数学建模（第3版） Giordano/59.00（附光盘）
微分方程与边界值问题（第5版） Zill/69.00
应用回归分析和其他多元方法（第3版） Kleinbaum/88.00
概率统计 Stone/89.00
多元数据分析 Lattin/69.00（附光盘）
数理统计与数据分析（第2版） Rice/78.00
随机过程导论 Kao/49.00
预测与时间序列（第3版） Bowerman/89.00
线性代数 Jain/49.00（附光盘）

复变函数及应用（第7版） Brown/42.00
实分析与复分析（第3版） Rudin/39.00
数学分析原理（第3版） Rudin/35.00
泛函分析（第2版） Rudin/42.00
复分析（第3版） Ahlfors/35.00
逼近论教程 Cheney/39.00
拓扑学（第2版） Munkres/59.00
数理金融初步（第2版） Ross/29.00
纯数学教程（第10版） Hardy/65.00
初等数论及其应用（第4版） Rosen/59.00
代数 Artin/59.00
实分析（第3版） Royden/45.00
流体动力学导论 Batchelor/66.00
组合数学教程（第2版） van Lint/58.00
曲线与曲面的微分几何 Carmo/49.00
三角级数（第3版） Zygmund/88.00
时间序列分析的小波方法 Percival /58.00
数学分析（第2版） Apostol/49.00
抽象代数基础教程（第2版） Rotman/49.00
分形分析 Kigami/35.00
图论 Tutte/45.00
计数组合学（卷1） Stanley/39.00

计数组合学（卷2） Stanley/59.00
同调代数导论 Weibel/45.00
数学简史 Katz/55.00
图论导引（第2版） West/59.00
复分析基础及工程应用（第3版） Saff/59.00
线性代数（第6版） Leon/59.00
初等数论及其应用（第5版） 罗森/69.00
贝叶斯方法 Leonard/45.00
调和分析导论（第3版） Katznelson/40.00
小波与小波变换导论 伯罗斯/29.00
组合数学（第4版） 布鲁迪/59.00

华章数学译丛

数学分析原理（原书第3版） Rudin/赵慈庚/28.00
金融数学 Stampfli/蔡明超/26.00
泛函分析（原书第2版） Rudin/刘培德/35.00
数学建模（原书第3版） Giordano/叶其孝/45.00
曲线与曲面的微分几何 do Carmo/田畴/39.00

复变函数及应用（原书第7版） Brown/邓冠铁/32.00
数理金融初步（原书第2版） Ross/陈典发/26.00
矩阵分析 Horn/杨奇/45.00
数学建模方法与分析（原书第2版） Meerschaert/刘来福/33.00
复分析（原书第3版） Ahlfors/赵志勇/33.00

重点大学计算机教材（相关部分）

数值方法 金一庆/25.00

计算机科学丛书（相关部分）

离散数学及其应用（原书第4版） Rosen/袁崇义/75.00
组合数学（原书第3版） Brualdi/冯舜玺/38.00
组合数学（原书第4版） Brualdi/冯舜玺/45.00

译 者 序

在工程技术、经济管理、科学研究和日常生活等诸多领域中，人们经常遇到的一类决策问题是：在一系列客观或主观限制条件下，寻求使所关注的某个或多个指标达到最大(或最小)的决策. 这种决策问题通常称为最优化问题，研究处理这类问题的数学方法称为最优化方法，它也是运筹学和管理科学中解决定量决策问题的基本方法. 在决策科学化的呼声日益高涨的今天，用最优化方法解决定量决策问题无疑是符合时代潮流和形势发展需要的.

线性规划是最优化问题领域中最简单、最基本和使用最广泛的方法，在许多决策领域都取得了巨大的成功. 学习一些线性规划的知识，无论对于科技工作者还是对于管理决策者来说，都是非常必要的.

受机械工业出版社委托，我们翻译了这本教材，正如原书作者在前言中所述，这本教材的主要特点在于起点较低，只假定读者具有很少的数学背景知识，这就给那些只具有中等数学知识背景的学生学习这门课程提供了一个机会.

原书中笔误和印刷错误较多，译者在翻译过程中根据原书网站上提供的勘误进行了修订并改正了新发现的错误.

本书第 1、2、3 章由姜启源翻译，第 4、5、6 章由张立平翻译，第 7、8 章由谢金星翻译，附录由秦添和谢金星翻译. 全书由谢金星统稿.

由于译者水平所限，不当之处欢迎读者不吝指正.

译　者

于清华园

前　言

为什么要写本书

这本教材是从一门高级的本科生数学课程演变而来的，课程的对象是具有不同数学基础和来自不同专业的学生，包括数学、计算机科学、统计学、工程科学、中等教育、精算学、计算机工程、理科和工商科等专业的学生. 有些是按五年制理科和工商科综合教学计划读工商管理硕士学位的学生，而有些是攻读博士学位的学生.

由于这门课程不需要诸如微积分、微分方程、抽象代数、拓扑学或数论等高深的数学理论，所以为那些只具有中等数学基础的学生学习一些实用而又重要的数学提供了一个机会. 考虑到这一点，只要可能，本书就尽量避免使用诸如向量空间、行列式、梯矩阵、极限和导数等高深或复杂的数学概念.

许多学生之所以选修这门课程，是因为线性规划在商业和其他领域有广泛的应用. 他们需要学习如何建立实际问题的公式，如何改写公式以便用具体的计算机软件求解，以及如何解释计算结果并应用于实际问题. 一旦计算机不能得到任何计算结果或者结果没有意义，他们应该能够对问题进行调整或者另外选择一个合适的软件.

在线性规划方面有许多优秀的教科书，但其中大多数需要很强的数学基础，只适于数学专业使用，或者只有知识超前的学生能够阅读，而包含的材料又大大超出一个学期的课程.

真正具有挑战性的是同时让高层次的学生和初学者在同一个课堂学习！在宾夕法尼亚州立大学，虽然线性代数是学习线性规划的先修课程，但课堂上有些学生解线性方程组存在困难. 另一方面，课堂上有些学生在数学或计算机科学方面却很强.

因此，我尝试不重蹈传统教材的老路，因为它们有点像民间故事中的金发姑娘拒绝食用的麦片粥一样——它们包含的题材要么太“冷”(内容过于平凡)，要么太“热”(需要严格的数学基础). 前一种情形会使许多学生感到厌烦；而在后一种情形会有一些内容使许多学生难于接受.

这本教材从入门开始介绍，只假定读者具有很少的数学基础. 因此，我给读者提供了一个机会，在学习线性规划的同时，在看到单纯形方法之前，首先了解线性代数和逻辑学中的相关工具. 有关逻辑的一节是线性规划的重要组成部分，虽然这一点经常被忽视. 在整本书中，我介绍了大量的例子和应用，并要求学生尝试不同难度的习题. 学生们很喜欢这种学习线性规划的方法，这可以从选修这

门课程的学生数量以及他们在学期末的评价表中给出的评价得到证实.

计算机应用的普遍性并没有消除对计算技巧的需求，但提高了逻辑技巧的相对重要性. 现在，如果你能通过手工计算得到圆周率的前100位数字，那只是出于好奇，而不能算是什么重要的结果，因为目前的计算机能够把圆周率的前 10^{10} 位数字算出来. 但是，从逻辑上看，是否有可能把圆周率的第 10^{100} 位数字算出来呢?

如何使用本书

这本教材是按照三个层次写的. 即使对于不了解线性代数和微积分的学生来说其中大部分内容也可以读懂. 对于程度更高的学生，本书给出了一些注释和习题. 在书后的附录中，给出了线性规划和数学规划其他方面发展的一般思想，为进一步研究提供指导. 附录中还给出了第1章至第8章中提到过的需要更强数学基础的一些主题的细节，并为高于典型的美国大学本科生水准的那部分学生提供经验和高深知识.

教材中给出了很多例子及其解答，所以我觉得没有必要给学生提供大量习题的答案. 即便如此，在本书的最后，我还是给出部分习题的答案[⊖]，包括那些比较棘手的习题. 习题的难度是不同的，但所有习题都可以用手工计算求解. 可以使用计算机，但在课堂上并不需要. 我没有提供用计算机求解具有很大优势的习题. 第1章第1节的习题除了可以检查对各种定义的理解外，还可以测试学生的数学基础.

致谢及参考资料

我的课堂讲稿经过了几年的演进，很多学生和阅卷评分者对讲稿的改进做出了贡献，他们指出其中的印刷错误和其他错误，并提出各种各样的问题. Prentice Hall出版社的审阅人和编辑也提了许多修正和改进意见.

我故意没有将本书与任何一个特定的软件包联系起来，因为我相信学生学习了本书的材料后，当他们面对一个好的软件包时，能够聪明地应用这些知识. 还有一个原因是，随着新软件包的出现以及计算机和操作系统的发展，任何一个特定的软件包都会很快过时.

但是，允许上课的学生使用他们喜欢的任何软硬件，即使在测验时也是如此. 能够求解线性规划问题的软件包包括Mathemetica、Maple、Excel等.

在因特网上有很多有关线性规划的软件，有的可以免费下载，有的可以在线使用. 因特网上也有很多关于线性规划的有用信息. 这里我列举一些网址，不过要记住，网上的变化是很快的：

⊖ 习题答案请参考华章网站(http://www.hzbook.com). ——编辑注

- http://carbon.cudenver.edu/hgreenbe/glossary/
  (数学规划词汇表)
- http://www.mathprog.org/
  (数学规划学会)
- http://iris.gmu.edu/asofer/siagopt.html
  (美国工业与应用数学学会最优化活动组)
- http://solon.cma.univie.ac.at/neum/glopt.html
  (全局最优化，维也纳)
- http://www.informs.org/Resources/
  (美国运筹学与管理科学学会)

在网络上以"线性规划"为关键词，可以搜索到很多网站. 有关线性规划的书也很多，在 2002 年 8 月 16 日，从网站 http://www.amazon.com 上检索到的有关"线性规划"的书多达 771 本.

也有很多杂志发表线性规划和非线性规划方面的文章. 在 2002 年 8 月 16 日，网站 http://www.informs.org/Resources/上列出运筹学方面的 36 种纸介质的杂志和 14 种在线杂志. 该网站还列出运筹学方面的 35 个学会.

Leon Vaserstein
vstein@math.psu.edu

目　录

第1章 引 言

1. 什么是线性规划

在经济领域中，用数学模型分析和优化经济问题的最早例子，也许是大约250年前由一位法国经济学家提出的. 在 François Quesnay (1694—1774) 写于1758年的《Tableau Économique》(经济表格)一书中，从分别考虑几个因素出发，解释了18世纪法国的地主、农民、工匠等角色的相互关系. 例如，书中有"关于国民现金收入的经济表格"和"每类财富的赢利和资本总额估计的经济表格".

早在1826年，19世纪的法国数学家傅里叶(Jean-Baptiste-Joseph Fourier, 1768—1830)关于线性不等式的研究表明，他对线性规划已经有所了解(见附录A.10节). 他还提出用单纯形算法求解线性逼近中的线性规划(见第8章). 19世纪后期，法国经济学家 L. Walras(1834—1910)的著作展示了他对线性规划的应用. 然而，除了像 Kantorovich 在1939年发表的专著《Mathematical Methods for Organization and Planning of Production》(生产的组织与计划的数学方法)这样少数值得注意的著作外，在第二次世界大战之前线性规划并没有引起太多注意.

计算机的发明和 George B. Dantzig 在1947年对单纯形算法的再发现这两件事情幸运地同步出现，对线性规划飞速的、突破式的发展，以及在经济、商业、工业工程、精算学、运筹学和对策论中的应用做出了贡献. 《纽约时报》显著地、充分地报导了线性规划的进展. 1970年 P. Samuelson (生于1915年) 获得诺贝尔经济学奖，1975年 L. Kantorovich(1912—1986)与 T. C. Koopmans(1910—1985)由于他们在线性规划中的工作获得诺贝尔经济学奖. 线性规划问题甚至进入到 Len Deighton 的充满悬念的侦探故事里，他在1966年出版的《Billion Dollar Brain》(十亿美元的大脑)中写道：

> "我不想打扰你，"Harvey 说，"但是你应该明白，实际上这一堆堆的线缆可以看作线性规划，它意味着不用一一考察所有的可能，他们就能预感到哪一个结果是正确的."

最优化问题有两种形式：最大化问题和最小化问题. 最大化问题是要在一个集合上使一个函数达到最大，最小化问题是要在一个集合上使一个函数达到最小.

在这两种情形下函数都是实值的，称为*目标函数*，集合称为*可行域*或*可行解的集合*. 求解一个最优化(最大化或最小化)问题通常指的是找可行域上的最优值——分别指最大值和最小值，以及一个最优解，即如果可能的话，如何(在何处)达到最优值. 除非另外说明，否则不需要求全部最优解，这与求解线性方程组要得到全部解不同.

最优值也称为极值，根据具体情况称极大值(maximum，max)和极小值(minimum，min)．全部最优解(极大值或极小值)集合称为最优域(optimality region)．

下面看一个简单的例子．

假设要求解以下的最优化问题：

$$\begin{cases}\text{maximize} & x\\ \text{subject to} & 2\leqslant x\leqslant 3.\end{cases}$$

显然其目标是给定了变量 x 取值的限制范围后，求它的最大值．因为这里的限制已经明确地表述为所考察的变量的函数，称为目标变量，所以这个问题不难解决，取最大值即可．于是得到正确结论：x 的最大值是 3，在 $x=3$ 处达到．

然而经常遇到的情况是，目标变量的取值范围隐含地由与目标变量有关的另一变量或其他一些变量的限制给出，这些变量称为决策变量或控制变量，它们在人们的控制之下：人们可以在给定的限制下随意地决定它们的数值．它们有别于作为优化问题输入的数据．目标函数总是决策变量的函数．目标函数有时也称为目标变量(objective variable)．

2

例如，在给定周长求什么样的矩形包围的面积最大这一问题中，目标变量是"面积"，决策变量是 l=矩形的长和 w=矩形的宽．一般来说，当目标变量以决策变量的函数形式给出时，我们用目标函数来表述要优化的函数．对于决策变量所加的限制称为问题的约束或限制．

于是，数学规划(mathematical programming，或 mathematical program)是个优化问题，其中目标函数是实变量(决策变量)的一个函数，可行域由附加在决策变量上的条件(约束)给出．一个可行解是满足问题中所有约束条件的决策变量取值的集合．

那么，什么是线性规划呢？线性规划是数学规划的一部分，它研究目标函数和约束条件都特别简单的优化(极值)问题．数学上，线性规划是以下形式的优化问题：在线性约束的有限集合上极大化或极小化一个仿射函数．与现今的一些用法不同，programming 一词不是指编写计算机程序，在本书中它回到军事计划，有点像指"详细计划"．

下面定义仿射函数和线性约束．本书中如不特别指明，数指"实数"，函数指"实值函数"．

定义 1.1 变量 x_1，…，x_n 的函数 f 称为线性型(linear form)，如果它可以表示为 $c_1x_1+\cdots+c_nx_n$，其中系数 c_i 是给定的实数(常数)．函数 f 称为仿射(affine)的，如果它是线性型和一个常数之和．

当然，不必一定将变量记作 x_i，系数记作 c_i，函数记作 f．例如，$g(x, y)=2x-a^2y$ 是两个变量 x，y(代替 x_1，x_2)的线性型，a 是给定的实数．注意，若 a 是一个变量，而 y 是给定的非零数，则 $f(x, a)=2x-a^2y$ 不是 x，a 的线性型

3

（见问题 1.2）．下面是两个变量 x，y 的 3 个仿射函数：$x-4y-3$，$y+2$，$x+y$.

问题 1.2　证明函数 $g(x, a)=2x-a^2y$ 不是 x，a 的线性型，其中 y 是给定的非零数.

解　反设 $g(x, a)=2x-a^2y$ 是 x，a 的线性型，即 $g(x, a)=2x-a^2y=c_1x+c_2a$，$c_1$，$c_2$ 是独立于 x 和 a 的系数，则 $g(1, 0)=2=c_1$，$g(0, 1)=-y=c_2$，于是 $g(x, a)=2x-a^2y=2x-ya$，得到 $a^2y=ya$ 对所有的 a 成立．取 $a=2$，则 $y=0$．但是根据假设 y 不能为 0，于是得到所希望的矛盾．■

线性函数(linear function)一词在一些教科书里指“线性型”，在另一些书里指“仿射函数”．线性泛函(linear functional)一词在线性规划中指“线性型”．

线性约束有 3 种形式：$=$，$\geqslant$，$\leqslant$．$=$ 形式的线性约束是常见的线性方程组，即形如

一个仿射函数＝一个仿射函数

的一组等式.

更常见的是标准形式

一个线性型＝一个常数.

例如，$x=2$，$x-y=0$，$5y=7$ 是以标准形式给出的两个变量 x，y 的 3 个方程，而 $2=x$，$x=y$，$3y+x+3=x-2y-4$ 是以不同形式给出的同样的方程.

另外两种形式的线性约束是形如

一个仿射函数(≤或≥)一个仿射函数

的不等式组.

通常它们写作

一个线性型(≤或≥)一个常数.

于是线性约束由用 3 个符号（$=$，$\geqslant$，$\leqslant$）之一连接的两个仿射函数（左边、右边各一个）构成，但 $x>0$ 那样的严格线性不等式不作为线性约束.

例 1.3

(i) $y=\sin 5$ 是对变量 y 的线性约束.

(ii) $x\geqslant 0$ 是对变量 x 的线性约束.

(iii) $2x+3y\leqslant 7$ 是对变量 x 和 y 的线性约束.

然而，注意

(iv) $y+\sin x=1$ 不是对变量 x 和 y 的线性约束，因为 $\sin x$ 对变量 x 不是线性型.　■

定义 1.4　线性规划(linear program 或 linear programming，LP)是在线性约束的有限集合上使一个仿射函数达到最大(或最小)的优化问题.

4

例如，下式是一个线性规划：

$$\begin{cases} \text{minimize} & f(x_1,\cdots,x_n) = c_1x_1 + \cdots + c_nx_n + d \\ \text{subject to} & \sum_{i=1}^{n} a_{ji}x_i \leqslant b_j \qquad j = 1,\cdots,m \\ & x_i \geqslant 0 \qquad i = 1,\cdots,n, \end{cases} \tag{1.5}$$

其中 m，n 是给定的自然数，d，c_i，b_j，a_{ji} 是常数，x_i 是决策(控制)变量(未知量). 称式(1.5)为线性规划的*典范型*.

定义 1.4 中约束的有限集合可以是空集，即约束的数目允许是零. 如果在优化问题中没有约束，我们称之为*无约束优化*. 注意，除非另外指出，否则不能忽略优化问题中给定的任何约束.

前面已经介绍线性型中不能有常数项，但是，常数项可以出现在线性规划的目标函数中. 于是按照定义，两个变量 x 和 y 的函数 $x-2y+3$ 不是线性型，而是一个仿射函数，它可以是线性规划的目标函数. 一些教科书中可能选择不同的定义.

可能出现一个优化问题甚至一个线性规划没有可行解的情况(见例 1.6). 这样的问题称为*不可行的*或*不相容的*. 也可能出现优化问题有可行解但是没有最优解的情况，如在 $x<1$ 条件下求 x 的最大值. 这个例子说明了在线性规划中不允许出现这种类型约束的原因.

一个优化问题称为*无界的*，如果在最大化问题中目标函数取无穷大，或在最小化问题中目标函数取无穷小(例 1.7). 在第 4 章将会看到，任何一个可行的线性规划或者有最优解，或者是无界的.

5

注意，可行解中可以有多个最优解，如例 1.8(或者一个也没有，如例 1.6). 然而，最优化(最大化或最小化)问题的最优值是唯一的(如果存在的话). 要是有两个不同的值，其中一个更好，那么另一个就不是最优的.

例 1.6 *一个不可行的 LP*

$$\begin{cases} \text{maximize} & 4x + 5y \\ \text{subject to} & 2x + y \leqslant 4 \\ & -2x - y \leqslant -5 \\ & x \geqslant 0, \quad y \geqslant 0. \end{cases}$$

注意，如果 x，y 满足约束 $2x+y\leqslant 4$，则乘以 -1 得到 $-2x-y\geqslant -4$，而第 2 个约束是 $-2x-y\leqslant -5$，显然这两个给定的约束互斥，因此没有可行解，这个线性规划是不可行的. ■

例 1.7 *一个无界的 LP*

$$\begin{cases} \text{maximize} & x - 2y \\ \text{subject to} & -3x + 2y \leqslant -2 \\ & -6x - 5y \leqslant -1 \\ & x \geqslant 0, \quad y \geqslant 0. \end{cases}$$

这个线性规划有一些可行解(如 $x=2/3$，$y=0$)，但是它们中没有一个是最优的. 对于任意实数 M，存在可行解 x，y，只要满足 $x-2y>M$. 例如 $x=2/3+M$，$y=0$. 从某种意义上说，有如此多的可行解，可是没有一个接近最优，这个线性规划是无界的. ■

例 1.8 有许多最优解的 LP

$$\begin{cases}\text{minimize} & x+y\\ \text{subject to} & x,y,z\geqslant 0.\end{cases}$$

在上述有 3 个变量 x，y，z 的问题中，最优解是 $x=y=0$，$z\geqslant 0$(任意非负实数)，最优值是 0. ■

例 1.9 有一个最优解的 LP

$$\begin{cases}\text{minimize} & x+y+z\\ \text{subject to} & x\geqslant -1, y\geqslant 2, z\geqslant 0.\end{cases}$$

6

在这个有 3 个变量 x，y，z 的线性规划中，最优解是 $x=-1$，$y=2$，$z=0$，最优值是 1. 注意，一个解必须包括所有含在该问题中的变量的数值. ■

例 1.10 一个非线性问题

$$\begin{cases}\text{minimize} & x^2+y^3+z^4\\ \text{subject to} & |x|\geqslant 1, |y|\leqslant 3.\end{cases}$$

这是有 3 个变量和 2 个约束的数学规划，它不是线性的，因为目标函数不是仿射的，并且约束不是线性的. (然而，第 2 个约束可以用两个线性约束代替，可行域是两部分的不相交并集，每部分可由 3 个线性约束给出.)毫无疑问，这个问题显然可以分解为 3 个独立的优化问题，每个问题只有一个变量. 于是恰有 2 个最优解，$x=\pm 1$，$y=-3$，$z=0$，最小值是 -26. ■

线性规划中所有的数都是实数. 事实上，很难想象来自商业和工业企业的一个线性规划会出现不是有理数的情况. 为什么? 试问，一个产品的价格能否用无理数(如$\sqrt{2}$)给出. 后面将会看到，为了求解带有理数数据的线性规划，我们不需要无理数. 而非线性问题就不是这种情况，如在解(非线性)方程 $x^2=2$ 时所得到的解.

为了进一步理解优化问题，试求解下面两个问题. 它们是线性规划吗?

问题 1.11

$$\begin{cases}\text{maximize} & x\\ \text{subject to} & 2\leqslant x\leqslant 3.\end{cases}$$

解 如前所述，最大值是 3(max=3)，在 $x=3$ 处达到. ■

在微积分中学过，函数的一阶导数的主要应用之一是，在临界点处求函数的极大值或极小值. 从问题 1.11 中可以看到，一年级微积分对于解线性规划是不够的. 如果想用一阶导数等于零的方法求线性型 $f(x)=x$ 在区间 $2\leqslant x\leqslant 3$ 上的最大值和最小值，可以看到，一阶导数是 1，不等于 0，目标函数在这个区间上

[7] 达到它的最大值 3 和最小值 2.

问题 1.12

$$\begin{cases}\text{maximize} & x+2y+z\\ \text{subject to} & x+y=1,\\ & z\geqslant 0.\end{cases}$$

解 当 $y\to\infty$，$x=1-y$，$z=0$ 时目标函数取任意大的值. 形式上可以写成 $\max=\infty$. 这是一个无界的线性规划. ■

问题 1.11 和问题 1.12 都是线性规划，因为目标函数及所有的约束都是线性的. 注意，最优化问题的最优值(最大值或最小值)是唯一的.（要是找到两个不同的值，其中一个更好，那么另一个不是最优的.）如果把符号 $-\infty$ 和 $+\infty$ 加入实数集合，作为最优值可以取的值，那么最优值总是存在的. 但是最优值可以用多种方法得到，即对于同样的最优值可以有许多最优解.

我们已经提到以下术语：线性型，仿射函数，线性约束，线性规划，目标函数，最优值，最优解，可行解. 试解释每个术语的含义.

注 我们已经介绍过，线性规划是数学规划的一部分. 至于数学规划，它又是运筹学(operations research 或 operational research)的一种工具，而运筹学是科学方法在有组织的军事、政府、商业、工业过程的管理与经营中的应用. 历史上，规划和运筹来自有计划的军事行动.

术语系统工程和管理科学几乎与运筹学有相同的含义，只是少强调和多强调一点人的因素而已. 作为运筹学的一部分，线性规划不仅仅是求解问题，而且涉及以下内容：

- 获取与处理决策所需要的数据.
- 列出问题与建立模型.
- 检验模型与解释求解结果.
- 将解用到决策中.
- 对决策进行控制.
- 对该过程的各个方面进行组织并使之相互联系.

[8] 本书不仅着重从数学上研究线性规划，而且还考虑把现实问题翻译成数学语言，把线性规划转换成不同形式，以及建立与对策论和统计学的联系.

线性规划与线性代数有何关系？线性代数主要是解线性方程组，在第 5 章将会看到，求解线性规划等价于对一组线性约束寻求可行解. 这样，从数学观点看，线性规划是更一般和更困难的问题.

注 除了数学规划以外，还有数学和计算机科学的一些其他领域，在那里最优化也起着重要作用. 例如，控制论和变分法与先前无法用有限个变量描述的最优化问题有关. 可行解可能是满足某些条件的函数. 我们

可以问什么是连接平面上两个定点的最短曲线，或者问什么是将任意规模的数据分类的最有效的方法．数学规划有时有助于这些问题的求解．

历史评注　这本书提到的数学家都是很著名的，他们的简历可以从百科全书、人物传记、历史书籍及万维网上找到．

Joseph Fourier，法国著名数学家、埃及古物学者和行政官员，以数学物理和工程中很重要的傅里叶级数而闻名．但他在线性逼近和数学规划中的工作不大为人所知．他是一个裁缝的儿子，有 11 个兄弟姐妹和 3 个同父异母(或同母异父)兄弟姐妹，在他 9 岁时母亲去世，第二年父亲去世．他接受了军事和宗教的教育，并卷入政治，他的生命几次陷于危险之中．

Leonid Kantorovich，对经济学有重要贡献的苏联数学家，在单纯形算法成功地被计算机实现并广泛地应用之前，他在美国几乎不为人所知．他 18 岁时得到数学博士学位．在几次数学讨论会以及与高级计划最优化有关的商务会议上，作者有幸在前苏联与 Kantorovich 会面．他在数学方面所作的许多工作之一，是在概率分布之间引入距离的概念，这一概念后来被其他数学家以不同的形式重新揭示，包括作者发现的距离(Vaserstein 距离)．对于类似于运输问题的题目，这个距离是最优值(见例 2.4 和第 6 章)．

9

练习

1～13. 说明以下各题是对还是错，并解释原因．

1. $1\leqslant 2$
2. $-10\leqslant -1$
3. $3\leqslant 3$
4. $-5/12\geqslant -3/7$
5. 对所有的数 x，y，$x^2+|y|\geqslant 0$.
6. 对所有的数 x，$3x\geqslant x$.
7. 对所有的数 x，$3x^3\geqslant 2x^2$.
8. 每个线性规划必须至少有一个线性约束.
9. 每个线性规划有一个最优解.
10. 线性规划中的每个变量必须是非负的.
11. 任何一个线性规划有唯一最优解.
12. 线性规划中约束的总数一定大于变量的个数.
13. 约束 $2x+5=6x-3$ 与一个 x 的线性方程等价.

14～17. 确定以下 x，y 的函数是否是线性型.

14. $2x$

15. $x+y+1$

16. $(\sin 1)x+e^x y$

17. $x \sin a+yz$

18～23. 以下对 x 是线性约束吗?

18. $x>2$

19. $|x| \leqslant 1$

20. $0=1$

21. $0 \geqslant 1$

22. $xy^2=3$

23. $ax=b$

24～26. 下面的陈述正确吗? 为什么?

24. $|x| \leqslant 1$ 等价于两个线性约束.

25. $|x| \geqslant 1$ 等价于两个线性约束.

26. 方程$(x-1)^2=0$ 等价于对 x 的一个线性约束.

27. 对 x 求解方程 $ax=b$，其中 a，b 是给定的数.

28～30. 对 x，y 求解以下 3 个线性方程组.

28. $\begin{cases} x+2y=3 \\ 5x+9y=4 \end{cases}$

[10]

29. $\begin{cases} x+2y=3 \\ 5x+10y=15 \end{cases}$

30. $\begin{cases} x+2y=3 \\ 3x+6y=0 \end{cases}$

31. minimize$(x+y)^2+(z+1)^2$，对所有的 x，y，z.

32. minimize $|x+2|+|x+3|$ subject to $|x| \leqslant 2.5$.

33. maximize $1/(1+x^2)$.

34. maximize $(x+y)^2+(z+1)^2$.

35. minimize $|x+y|+(z+1)^6+(x-y+z)^2$.

36～43. 以下是两个变量 x，y 的线性型吗?（回答是或否）

36. $2x+3y$

37. $2x+3y=1$

38. $x+y^2$

39. xy

40. y

41. 0

42. $(x+1)^2+2y-x^2-1$

43. x/y

44～49. 以下对两个变量 x，y 的约束是线性的吗？

44. $xy=0$

45. $x=0$

46. $x<0$

47. $x+y=0$

48. x 为整数

49. $x\geqslant 1$ 或 $y=0$

50～57. 以下对两个变量 x，y 是线性约束吗？（回答是或否）

50. $x+2y$

51. $x\geqslant 1$

52. $0=0$

53. $0=1$

54. $x+y\leqslant 0$

55. $x^2=2$

56. $x\geqslant 0$

57. $xy=0$

58. 说明两个变量 x，y 的任意一个线性型 $f(x, y)$都有以下两条性质：

- （比例性）对于每个数 a，$f(ax, ay)=af(x, y)$
- （可加性）$f(x_1+x_2, y_1+y_2)=f(x_1, y_1)+f(x_2, y_2)$

59. 反之，说明具有这两条性质的任意两变量 x，y 的函数 $f(x, y)$是线性的.

60. minimize $|x|+(x-2y)^2+\sin z+2^u+\log(v+101)$ subject to $|x|$，$|y|$，$|z|$，$|u|$，$|v|\leqslant 100$. 有多少最优解？

提示：对目标函数分别求每一项的极小值.

11

2. 线性规划的例子

我们曾在第 1 节中提到线性规划出自商业和工业，下面两节给出一些能够表述为线性规划的现实生活中的例子.

例 2.1　食谱问题

在当前这个普遍具有健康意识的时代，许多人在分析食物的营养成分. 让我们看看这如何构成一个线性规划.

选择不同食物的组合作为食谱的一般想法是，以最小费用满足对基本营养的需求.

当然，这类实际问题是相当复杂的，我们必须从营养学家那里知道什么是基本营养需求(可能因人而异). 另外，为了保持多样性以避免对营养食物的厌倦，应该考虑一个较长的可选择食物的清单. 下面的例子做了很多的简化.

按照营养学家的建议，一个人一天对蛋白质、维生素 A 和钙的需求如下：50 g 蛋白质、4 000 IU(国际单位)维生素 A 和 1 000 mg 钙. 为简单起见，我们只考虑以下食物构成的食谱：苹果(生的，带皮)、香蕉(生的)、胡萝卜(生的)、枣(本国的，野生的，去核的，切碎的)和鸡蛋(整个的，生的，新鲜的). 如果可能的话，确定每种食物的用量，以最小费用满足推荐的定额(Recommended Dietary Allowance，RDA).

食物	单　位	蛋白质(g)	维生素 A(IU)	钙(mg)
苹果	中等大小一个(138 g)	0.3	73	9.6
香蕉	中等大小一个(118 g)	1.2	96	7
胡萝卜	中等大小一个(72 g)	0.7	20 253	19
枣	一杯(178 g)	3.5	890	57
鸡蛋	中等大小一个(44 g)	5.5	279	22

12

由于目的是以最小费用满足 RDA，所以需要搜集这些食物的价格：

食物	价格(美分)
一个苹果	10
一个香蕉	15
一个胡萝卜	5
一杯枣	60
一个鸡蛋	8

现在用这些数据建立一个线性规划. 变量 a，b，c，d，e 表示列入食谱的 5 种食物的数量，要最小化的目标函数是总费用函数(以美分计)，

$$C = 10a + 15b + 5c + 60d + 8e,$$

其中系数为 5 种食物每单位的价格.

约束是什么？显然，

$$a,b,c,d,e \geqslant 0. \tag{i}$$

它们称为非负约束(nonnegativity constraint).

其次，为保证满足对蛋白质、维生素 A 和钙的最小日需求量，必须有

$$\begin{cases} 0.3a + 1.2b + 0.7c + 3.5d + 5.5e \geqslant 50 \\ 73a + 96b + 20\,253c + 890d + 279e \geqslant 4\,000 \\ 9.6a + 7b + 19c + 57d + 22e \geqslant 1\,000, \end{cases} \tag{ii}$$

其中，对于第 1 个约束，第 1 项 $0.3a$ 表示一个苹果中蛋白质的含量(g)乘以食谱中苹果的数量，第 2 项 $1.2b$ 表示一个香蕉中蛋白质的含量(g)乘以食谱中香蕉的数量，依此类推.

注意，第 1 个约束里的各项以 g 计，第 2 个约束里的各项以 IU 计，第 3 个约束里的各项以 mg 计.

我们知道，求解这个线性规划需要寻求一个最优值和一个最优解．让我们用反复试验的办法找这个问题的解．因为胡萝卜便宜，先看一个全是胡萝卜的食谱，即 a，b，d，e 等于 0.(ⅱ)的 3 个约束简化为

$$\begin{cases} 0.7c \geqslant 50 \\ 20\,253c \geqslant 4\,000 \\ 19c \geqslant 1\,000. \end{cases} \tag{iii}$$

这样，当 $c=500/7$(约 71 个胡萝卜)时食谱中蛋白质的需求刚好满足，而维生素 A 和钙的需求大体上超过．因为每个胡萝卜 5 美分，所以这个食谱的费用是 25/7 美元．能够更好些吗？能否找到除胡萝卜外由另外食物构成的食谱，而每天的费用比 $25/7 \approx 3.57$ 美元便宜？(每天吃 71 个胡萝卜的人希望如此！)

因为鸡蛋中的蛋白质特别丰富，在蛋白质需求量的约束中 e 的系数相当大，这启示我们，如果在菜单中加入几个鸡蛋，既满足了我们自己设置的营养需求，又避免了食谱的单调．尝试用增加鸡蛋数量 e 的办法减少胡萝卜的数量 c．保持 a，b 和 d 等于 0，现在 3 个约束为

$$\begin{cases} 0.7c + 5.5e \geqslant 50 & \text{(蛋白质数量 g)} \\ 20\,253c + 279e \geqslant 4\,000 & \text{(维生素 A 数量 IU)} \\ 19c + 22e \geqslant 1\,000 & \text{(钙数量 mg).} \end{cases} \tag{iv}$$

容易看出，当 $c=50$，$e=3$ 时这些约束满足．由于每个胡萝卜 5 美分，每个鸡蛋 8 美分，这个新食谱的费用是 2.74 美元．瞧！从每天 71 个胡萝卜的食谱到 50 个胡萝卜和 3 个鸡蛋构成的新食谱，不仅受欢迎而且便宜很多．还能够做得更好吗？

虽然反复试验的办法帮助我们做了一些分析，但是并没有对诸如在食谱中怎样考虑 5 种食物的其他组合以进一步降低费用给出指导原则．在我们学习了单纯形算法(第 5 章)以后再回到这个例子．

注　下面是关于食谱问题的一些值得思考的事情.

(a)每种食物成分都有自己的单位，为避免混淆起见应关注单位的名称，也要指明费用目标函数的单位(可以表示为美分、美元、千美元等).

(b)答案中允许鸡蛋的数目以分数(如 1/2)给出吗？在线性规划的形式解中回答是肯定的(线性规划的可分性假设)．这样做有意义吗？它取决于讨论的问题．有时我们要求某些变量是整数，增加这些非线性约束将把一个线性规划变成整数线性规划.

(c)如何解食谱问题？我们将在第 5 章讨论单纯形算法时回答这个问题．前面已经用反复试验的办法进行了几次迭代，虽然得到了两个可行解，但是仍然不知道最优解.

(d)因为苹果的大小不同，要用重量单位而不是用“个”来度量它吗？视具体情况而定. 选择的单位取决于现实生活. 如果在便利店买苹果，大小都差不多，你按“个”付款，那么在这种情况下“个”就是合适的单位；如果在农场里用一蒲式耳[⊖]的容器来买，那么蒲式耳(bushel)是合适的单位. 对于食物成分的变化只在不同的重量单位间进行转换不是完美的解决办法，虽然有时也有用. 现实生活中数量经常是不确切的(即使它们都知道)，当你甚至无法保证知道明天的价格时，能够计划下个月的食谱吗？另外有人做这件事，有时你也必须做. ■

让我们提出目标是最小化费用的另一个问题.

例 2.2 混合问题

不同国家的许多硬币都是用白铜(75%铜，25%镍)做成的. 假设生产硬币的 [15] 4 种可用的合金(碎金属片)A，B，C，D 所含铜和镍的百分比如下表：

合金	A	B	C	D
% 铜	90	80	70	60
% 镍	10	20	30	40
美元/lb	1.2	1.4	1.7	1.9

每种合金的费用以美元/lb 形式在表的最后一行给出.

注意 4 种合金没有一种含有所要求的铜和镍的百分比，我们的目的是把这些合金组合成一种新的混合物，它含铜和镍的百分比符合白铜的要求，并且费用最低. 这个问题适合表示为线性规划.

设 a，b，c，d 是合成 1lb 新混合物所需的合金 A，B，C，D 的数量(以 lb 计)，有

$$a,b,c,d \geqslant 0. \qquad (\text{i})$$

由于新混合物只由这 4 种合金组成，所以

$$a+b+c+d=1. \qquad (\text{ii})$$

新混合物的合金条件为

$$\begin{cases} 0.9a+0.8b+0.7c+0.6d=0.75 \\ 0.1a+0.2b+0.3c+0.4d=0.25. \end{cases} \qquad (\text{iii})$$

例如，第 1 个等式的意思是，合金 A 数量的 90%，加合金 B 数量的 80%，加合金 C 数量的 70%，加合金 D 数量的 60%，等于 1lb 新混合物所需的 75%的铜. 类似地，第 2 个等式给出了新混合物所需镍的数量.

注意到以上的约束，求以下费用函数的最小值：

$$C=1.2a+1.4b+1.7c+1.9d.$$

⊖ 蒲式耳(容量等于 8 加仑，加仑的单位符号为 USgal)，1USgal=3.785 41dm^3.

在这个问题中除了(ⅰ)式外所有的约束都是等式．事实上，有3个线性方程和4个未知数，然而3个方程不是独立的，例如可由(ⅲ)式方程组求和得到(ⅱ)式，于是(ⅱ)式是多余的． ■ [16]

一般地，一个约束称为多余的(redundant)，如果它可以从这个规划问题的其他约束得到．由于多余的约束对于线性规划的解没有贡献新的信息，它可以删除而不影响可行集合．

例 2.3 制造问题

现在叙述一个规划问题，要最大化的目标函数是利润．一家工厂生产3种产品：P1，P2，P3．每种产品的度量单位都是放置产品的标准尺寸的箱子．每箱P1，P2，P3的利润分别是2美元，3美元和7美元．记 x_1，x_2，x_3 分别为P1，P2，P3的数量，则要最大化的利润函数是

$$P = 2x_1 + 3x_2 + 7x_3.$$

要利用的5种资源是原材料R1和R2、生产时间、生产场所面积及机器运行时间．现有1 200 lb的R1，300 lb的R2，40 h生产时间，8 000 m^2 生产场所面积和8 h机器运行时间．

每箱产品所需的每种资源的数量由下表给出(包括前面的数据)：

资　源	单位	P1	P2	P3	现有量
R1	lb	40	20	60	1 200
R2	lb	4	1	6	300
生产时间	h	0.2	0.7	2	40
生产场所面积	m^2	100	100	800	8 000
机器运行时间	h	0.1	0.3	0.8	8
利润	美元	2	3	7	→ 最大利润

如表中所见，生产1箱P1需要40 lb R1，4 lb R2，0.2 h生产时间，100 m^2 生产场所面积和0.1 h机器运行时间．生产1箱P2，P3需要的资源也由此表可知．

约束为

$$x_1, x_2, x_3 \geqslant 0, \qquad (\text{i})$$ [17]

$$\begin{cases} 40x_1 + 20x_2 + 60x_3 \leqslant 1\,200 & (\text{R1 数量 lb}) \\ 4x_1 + x_2 + 6x_3 \leqslant 300 & (\text{R2 数量 lb}) \\ 0.2x_1 + 0.7x_2 + 2x_3 \leqslant 40 & (\text{生产时间 h}) \\ 100x_1 + 100x_2 + 800x_3 \leqslant 8\,000 & (\text{生产场所面积 m}^2) \\ 0.1x_1 + 0.3x_2 + 0.8x_3 \leqslant 8 & (\text{机器运行时间 h}). \end{cases} \qquad (\text{ii})$$

显然最优解的第一个自然近似是只生产P3，因为它每箱的利润大于P1，P2．这意味着应令 $x_1=0$，$x_2=0$．于是约束(ⅱ)式变为

$$\begin{cases} 60x_3 \leqslant 1\,200 \\ 6x_3 \leqslant 300 \\ 2x_3 \leqslant 40 \\ 800x_3 \leqslant 8\,000 \\ 0.8x_3 \leqslant 8. \end{cases} \qquad \text{(iii)}$$

上述不等式的一个解是 $x_3=10$，利润 P=70. 然而当我们算出生产 10 箱 P3 所用的资源时可以看到，600 lb R1 剩下没用，多于一半的 R2 没用，一半生产时间浪费了. 注意，最大的利润总是伴随着资源的最优利用，以达到投资的最好回报. 虽然这个初步近似是可行的，但是我们猜测，对于这个问题它不是最优的. 后面将会看到，最优解是 $x_1=20$，$x_2=20$，$x_3=0$，利润 max=100. ■

例 2.4 运输问题

制造商经常关注的另一个问题是产品的运输费用. 我们考察下面的假想情况，试着建立一个线性规划.

一家小型机器制造商在亚特兰大、巴尔的摩和芝加哥备有仓库，亚特兰大的仓库里有 50 台机器，巴尔的摩的仓库里有 30 台机器，芝加哥的仓库里有 50 台机器. 在底特律、尤金、费尔维尤、格鲁弗城及休斯敦设有零售店，它们分别至少需要 25，10，20，30，15 台机器. 显然，制造商要从 3 个仓库把机器运送到 5 个地方的商店，并且尽量以最便宜的方式到达. 上述问题给出了线性规划的一个理想的背景——最小化运输费用. 为解决这个问题，我们需要知道从每一个仓库运输一台机器到每一个商店所需的费用，这由运费表给出.

18

	1. 底特律	2. 尤金	3. 费尔维尤	4. 格鲁弗城	5. 休斯敦
1. 亚特兰大	55	30	40	50	40
2. 巴尔的摩	35	30	100	45	60
3. 芝加哥	40	60	95	35	30

这样，从巴尔的摩到尤金运输一个单位的产品要花费 30 美元，从芝加哥到费尔维尤要花费 95 美元，等等.

为了把这个问题表述成线性规划，引入变量表示从每一个仓库到每一个商店运输产品的数量. 我们依字母顺序将仓库编号，并类似地将商店编号. 设 x_{ij} ($1\leqslant i\leqslant 3$, $1\leqslant j\leqslant 5$) 为从仓库 i 到商店 j 运输机器的数量，这给出 15 个未知数. 目标函数(要最小化的量)是运输费用，由下式给出：

$$\begin{aligned} C = &\ 55x_{11}+30x_{12}+40x_{13}+50x_{14}+40x_{15} \\ &+35x_{21}+30x_{22}+100x_{23}+45x_{24}+60x_{25} \\ &+40x_{31}+60x_{32}+95x_{33}+35x_{34}+30x_{35} \end{aligned}$$

其中 $55x_{11}$ 是从亚特兰大的仓库运输一台机器到底特律的商店所需的费用乘以要运输的机器数量，等等.

约束是什么？首先，运输的机器数量为负数没有意义，所以 15 个变量满足条件

$$x_{ij} \geqslant 0, 1 \leqslant i \leqslant 3, 1 \leqslant j \leqslant 5 \tag{i}$$

其次，从仓库 i 运走机器的数量不能超过它的库存，故有

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} + x_{15} \leqslant 50 \\ x_{21} + x_{22} + x_{23} + x_{24} + x_{25} \leqslant 30 \\ x_{31} + x_{32} + x_{33} + x_{34} + x_{35} \leqslant 50, \end{cases} \tag{ii}$$

19

再次，考虑到每个零售店所需的机器数量，得到下面 5 个约束：

$$\begin{cases} x_{11} + x_{21} + x_{31} \geqslant 25 \\ x_{12} + x_{22} + x_{32} \geqslant 10 \\ x_{13} + x_{23} + x_{33} \geqslant 20 \\ x_{14} + x_{24} + x_{34} \geqslant 30 \\ x_{15} + x_{25} + x_{35} \geqslant 15. \end{cases} \tag{iii}$$

问题已经给出，以后将会研究求解这类问题的有效方法，现在我们仍然用反复试验的办法摸索着进行.

从地理位置和财政开支上说，似乎合理的办法是，从亚特兰大的仓库运输机器到尤金和费尔维尤的商店，而格鲁弗城和休斯敦对芝加哥是有吸引力的市场. 于是

$$\begin{cases} x_{12} = 10, x_{22} = x_{32} = 0 \\ x_{13} = 20, x_{23} = x_{33} = 0 \\ x_{34} = 30, x_{35} = 15. \end{cases}$$

现在我们确定了 15 个变量中的 8 个. 继续下去，由于尤金和费尔维尤的商店从亚特兰大的仓库得到需要的机器，格鲁弗城和休斯敦的商店从芝加哥的仓库得到需要的机器，使得再从其他的仓库运输更多的机器变得毫无意义，因为这样只会增加运输费用. 所以

$$x_{14} = x_{15} = x_{24} = x_{25} = 0.$$

只剩下 3 个变量 x_{11}，x_{21}，x_{31} 待定. 利用约束可以确定

$$x_{21} = 25, x_{11} = x_{31} = 0. \quad ■$$

关于该可行解的评注

(a)注意在可行解中要运到底特律的机器数量是 25，要运到尤金的机器数量是 10，等等. 如果运输比每个商店的需求更多的机器，只会增加费用. 另一方面，我们并未用到仓库供应的全部机器，因为供应是

[20] 大于需求的.

(b)在现实生活中总的运输费用并不都是用线性型表示，例如大批量运输可以享受到折扣. 求解非线性问题通常比线性问题困难得多.

例 2.5 工作指派问题

假设一位生产管理者需要指派 n 个工人去做 n 项工作，如果每个工人完成每项工作的技术水平和效率一样，那么工作可以以任意方式分配. 然而正如我们所知，这种情况很少见. 如果每个工人都按照他完成每项工作的时间来评估，这个时间(以小时计)表示为大于或等于零的数. 显然，目标是以总时间尽可能小的办法指派工人的工作. 设 c_{ij} 是工人 i 完成工作 j 需要的时间，于是这些时间自然地写成一个表的形式，如取 $n=3$，时间由下表给出：

	a	b	c
A	10	70	40
B	20	60	10
C	10	20	90

即若指派工人 A 做工作 a，工人 B 做工作 b，工人 C 做工作 c，则总时间为 $10+60+90=160$. 这不是最小的，因为若 A 做 b，B 做 c，C 做 a，则总时间等于 $70+10+10=90$. 对于 $n=3$，可能的工作指派方式的总数是 $3\times 2=6$，这可从包含在下表的信息中看到：

指派			总时间
Aa	Bb	Cc	160
Aa	Bc	Cb	40
Ab	Ba	Cc	180
Ab	Bc	Ca	90
Ac	Ba	Cb	80
Ac	Bb	Ca	110

[21] 由此表可知，总时间的最小值是 40. 结论是管理者聪明的做法是指派工人 A 做 a，工人 B 做 c，工人 C 做 b.

一般来说，这种选择方法不好，可能的工作指派方式的总数是 $n!=n\times(n-1)\times\cdots\times 2\times 1$，即使对于中等大小的 n，这也是一个天文数字，如对 $n=70$，

$$n! = 11\,978\,571\,669\,969\,891\,796\,072\,783\,721\,689\,098\,736\,458\,938\,142\,546\,425\,857\,555\,362\,864\,628\,009\,582\,789\,845\,319\,680\,000\,000\,000\,000\,000.$$

人们估计，如果 Sun 工作站计算机从 Big Bang(宇宙大爆炸)时开始用检查所有可能的工作指派方式解这个问题，那么直到现在它还没有完成这项任务.

工作指派问题能够表示为一个线性规划，虽然这不是显而易见的，但是这样它就能用单纯形算法求解. 当 $n=70$ 时需要几秒钟. 在讨论了单纯形算法之后，第 7 章我们将回到这里并研究相关问题. ■

用一段引语作为这一节的结束. 1980 年 Eugene Lawler 写道：

> "线性规划"可用于资源分配、生产计划制定、人力调度、投资安排以及确定市场(及军事)策略等，在今天的工业社会里线性规划的多功能性和对经济的影响的确是巨大的.

练习

1. 假设为了使食谱平衡你只需要蛋白质和维生素 A，B_1，C，B_6，B_{12}，但只允许你吃各种谷物. 到一家食品杂货店去选择 10 种不同箱子的谷物. 写出每种谷物含上面这些成分的每日推荐量(RDA)的百分比，以及它们的价格，然后写出你的食谱问题，注明单位、日期和商店. 如果需要的话，你可以选择你的性别、年龄和卡路里摄入量来确定你的 RDA. 如果从箱子上和其他地方无法找到你的蛋白质 RDA，就取为 50g. 记下商店的名字和日期. 你可以从因特网上的一家商店收集数据.

2. 解例 2.2 的混合问题. 提示：注意例 2.2 的(ⅲ)式是 4 个未知数、2 个线性方程的方程组，能够消去两个变量，得到两个变量的问题可以用图解法求解(见第 3 节). 22

3. 增加条件 $x_3=0$，解例 2.3 的线性规划. 提示：利用图解法(见第 3 节).

4. 你有 100 个 25 美分的硬币(quarter)和 90 个 10 美分的硬币(dime)，没有其他钱币. 你必须支付给定的款额 C，不找零给你，而你也不愿支付过多的款项. 用数学形式表述这个优化问题(即借助于决策变量、目标函数和约束条件). 这个问题是线性的吗? 对 $C=15$ 美分，$C=1.02$ 美元，$C=100$ 美元求解这个问题.

5. 有一个长 100 的线环，你想围成一个面积最大的矩形. 从数学上提出这个问题并求解.

6. 例 2.5 有一种有趣的变形. 假设生产管理者设计了一种系统，由此可以按照每个雇员完成一件特定工作的好坏给他确定一个数量等级，然后管理者想以雇员们等级之和尽可能地大来指派哪个人做哪项工作，这样公司的效率就会尽可能高. 你的公司很小，只对 4 名雇员根据完成要指派工作的好坏确定了等级. 利用下面的等级矩阵求解这个最大化问题：

	a	b	c	d
A	10	70	40	55
B	20	60	10	67
C	10	20	90	43
D	15	37	89	23

7. 上述问题的另一种解释是**匹配问题**(matching problem). 你想把 n 个男孩和 n 个女孩配成 n 对夫妻，目标是得到最大的幸福(或最小的悲伤). 在下面的数据中，正数是预料一对夫妻因幸福的婚姻而付给你的钱(以美元计)，负数是一对夫妻因不幸福的婚姻而从你那里拿走的钱，且 $n=4$：

23

	a	b	c	d
A	1	−2	2	0
B	2	0	1	1
C	3	1	0	−1
D	−1	0	1	2

求解这个问题. *提示*：如果你没有更快的办法就检查全部 4! =24 种匹配.

8～10. 求解下面的匹配问题，表中的数是预料每对夫妻一起幸福度过的年数，要使这些年数的总和最大. 如果你不能在一个合理的时间段内解决这个问题，就尝试寻求尽可能好的可行解. 数据不是取自现实生活.

8.

	a	b	c	d	e
A	8	2	9	0	0
B	2	9	1	1	3
C	3	1	7	1	1
D	1	6	1	2	9
E	8	8	1	9	1

9.

	a	b	c	d	e	f	g
A	8	2	9	0	3	8	7
B	2	0	1	1	3	7	9
C	3	1	1	1	1	6	9
D	1	0	1	2	9	5	8
E	8	8	1	1	1	5	7
F	1	6	1	9	9	5	8
G	6	6	6	5	1	5	4

10.

	a	b	c	d	e	f	g
A	8	2	5	0	0	1	7
B	2	9	1	1	3	7	5
C	3	1	7	1	1	6	9
D	1	6	1	2	1	5	8
E	8	8	1	0	1	5	7
F	0	6	1	9	1	5	8
G	6	6	1	5	1	5	4

11. 从给定的数 c_1，…，c_n 中寻找最大的数. 把这个问题表述成线性规划.

12. 给定 3 个不同的数 a，b，c，从中找出中位数(即给出的数中不是最大也不是最小的那个数 x). 把这个问题表述成线性规划. 提示：这是个困难问题，但是将在第 8 章解决.

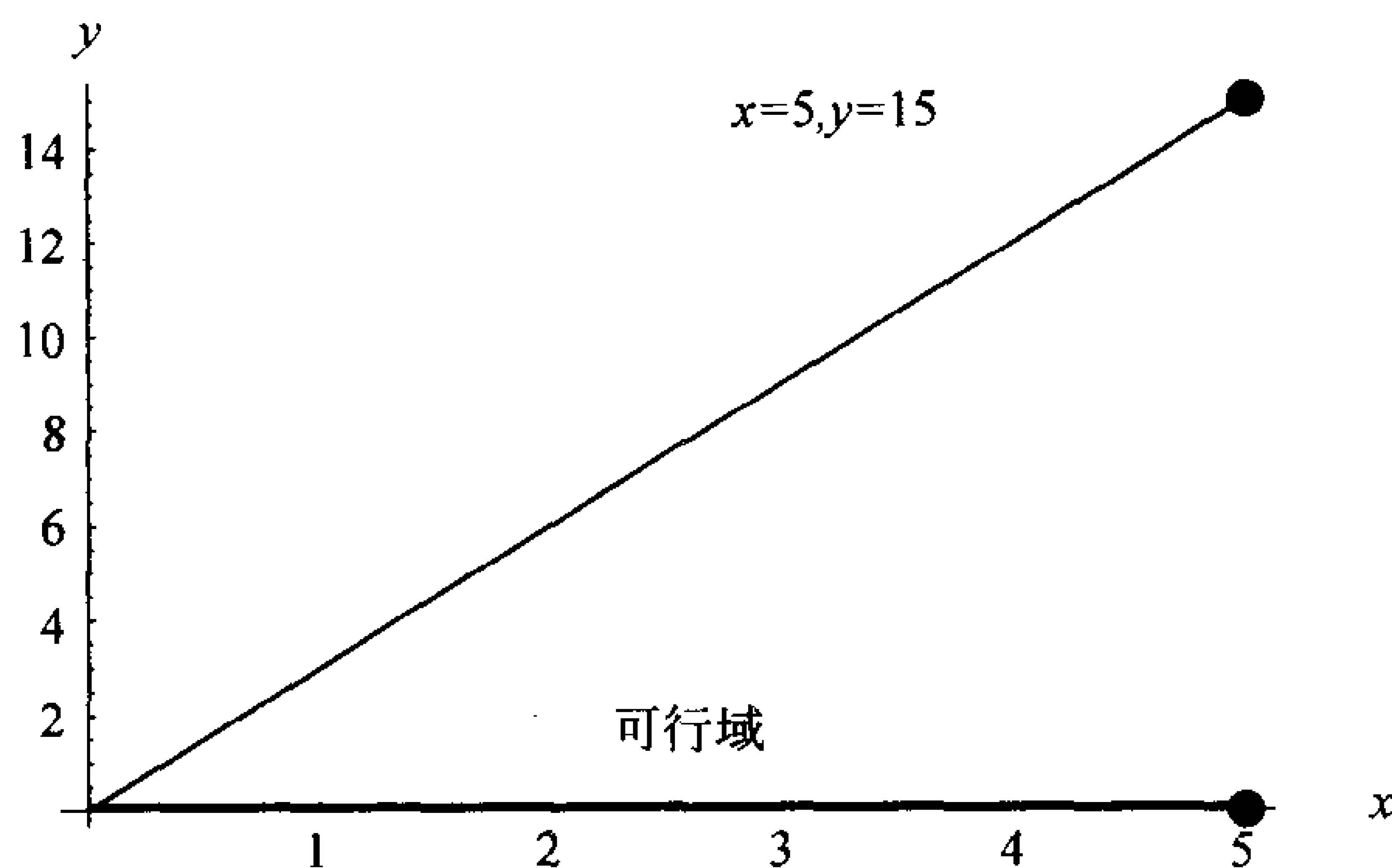
24

3. 图解法

例 3.1 考察下面这个简单的线性规划：

$$\begin{cases}\text{maximize} & y = 3x \\ \text{subject to} & 0 \leqslant x \leqslant 5.\end{cases}$$

观察下面的图，使这个函数达到最优的 x 的值是明显的：



一般来说，若决策变量的个数不超过 2，就可以用平面上的图求解优化问题. 但在高维的情况下要困难得多.

对于含一个决策变量 x 的线性规划，目标函数 y 在(x, y)平面上的图形是一条直线，可行域是 x 轴上的一个集合. 一般地，一个变量 x 的任何一组有限的线性约束，等价于单个的线性约束，或者一组两个的线性约束. 为了说明这一

点，用下面 5 种约束类型之一代替每一个约束：$0=0$，$0=1$，$x\geqslant b$，$x\leqslant b$，$x=b$. 因此，一个变量 x 的线性约束的可行域具有下面 6 种形式之一：

- 整条直线.
- 从 $-\infty$ 到数 c 的射线 $x\leqslant c$.
- 从数 c 到 $+\infty$ 的射线 $x\geqslant c$.
- 带端点 a，b ($a<b$) 的闭区间 $a\leqslant x\leqslant b$.
- 一个点 $x=c$.
- 空集(即没有可行解). ■

[25]

将检验下面的论据作为练习留给读者：

论据 3.2 若目标函数是常数函数 $y=c$，则最优值是 $y=c$，它在可行域内任意点 x 达到.

论据 3.3 设目标函数形式为 $y=\alpha x$，其中 α 是非零实数.

1. 如果可行域是整条直线，则线性规划没有最优解.

2. 如果可行域是射线 $x\leqslant c$(或相反地，$x\geqslant c$)，则线性规划有一个最优解当且仅当它是一个最大化问题(或相反地，最小化问题)，最优值 $y=\alpha c$ 在点 $x=c$ 达到.

3. 如果可行域是闭区间 $a\leqslant x\leqslant b(a\leqslant b)$，则最优值在端点 a 或 b 达到.

4. 如果可行域是空集，则没有最优解.

当决策变量个数为 2 时，为了寻找线性规划所有的可行解和最优解，可以在笛卡儿平面上作图，称为线性规划的图解法(graphing method). 由此得到的几何上的解释可以用于任意多个变量的情况.

有时变量个数很多的问题可以缩减为较少变量个数的问题. 如在例 2.2 中变量个数可以从 4 减为 2(不计目标变量 C).

下面从变量个数为 1 的例子开始.

例 3.4

$$\begin{cases}\text{minimize} & y = 0.3x \\ \text{subject to} & 2x \leqslant 50, \\ & 3x \leqslant 120, \\ & -5x \geqslant -250, \\ & 0.5x \geqslant -3.\end{cases}$$

这个线性规划可以作图如下. 每个不等式给出 x 轴上的一条射线：不等式 $2x\leqslant 50$ 表示射线$(-\infty, 25]$，约束 $3x\leqslant 120$ 表示射线$(-\infty, 40]$，约束 $-5x\geqslant -250$ 表示射线$(-\infty, 50]$. 例 3.4 的第 4 个约束表示什么射线？由于每一个约束所起的作用都要考虑，可行域必定是这些射线的公共部分或交集. 由此可知可行域是区间 $-6\leqslant x\leqslant 25$，即所有可行解的集合.

[26]

图 3.5 给出可行域上目标函数的图形.

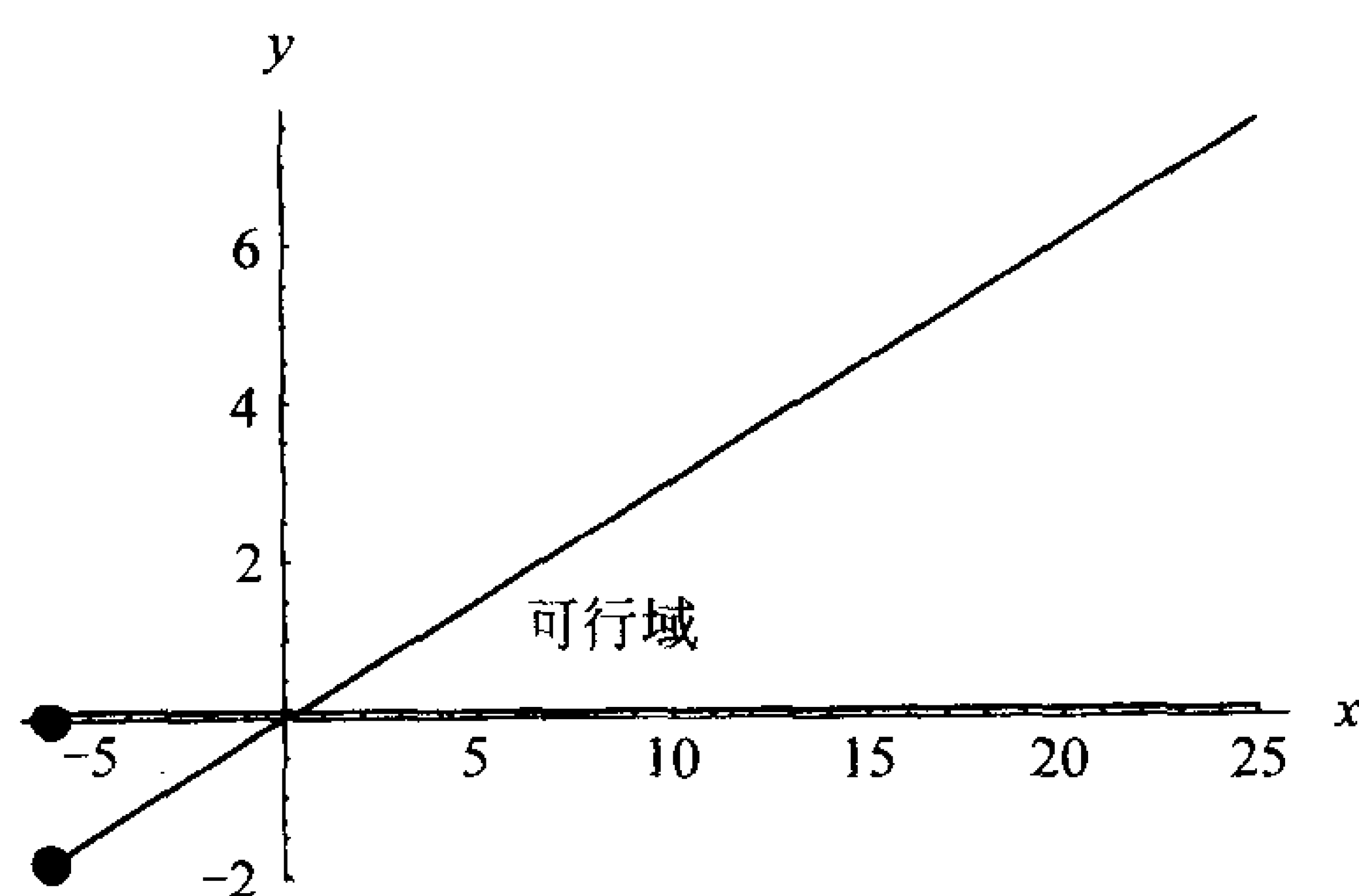


图　3.5

显然，这个问题的目标函数的最小值是-1.8，在$x=-6$处达到，是唯一的最优解．答案：$x=-6$，$\min=-1.8$．■

最优解集合也具有上面6种形式之一，是可行域的子集．当可行域是一个区间时，最优解是区间的端点之一．当且仅当目标函数不是常数时，有唯一最优解（它必定是一个端点）．

现在考察有两个变量的线性规划．

例 3.6

$$\begin{cases} \text{maximize} & f = x + 9y \\ \text{subject to} & x \geqslant 0, y \geqslant 0, \\ & x - y \leqslant 3, \\ & x - 3y \geqslant -5, \\ & 5x + 7y \leqslant 35. \end{cases}$$

首先在(x, y)平面上画出可行域F，它由满足线性约束的所有点(x, y)构成．由于每个约束都是线性不等式，而满足一个不等式的图形是半平面，所以可行域F是所有这些半平面的交集（图3.7）． 27

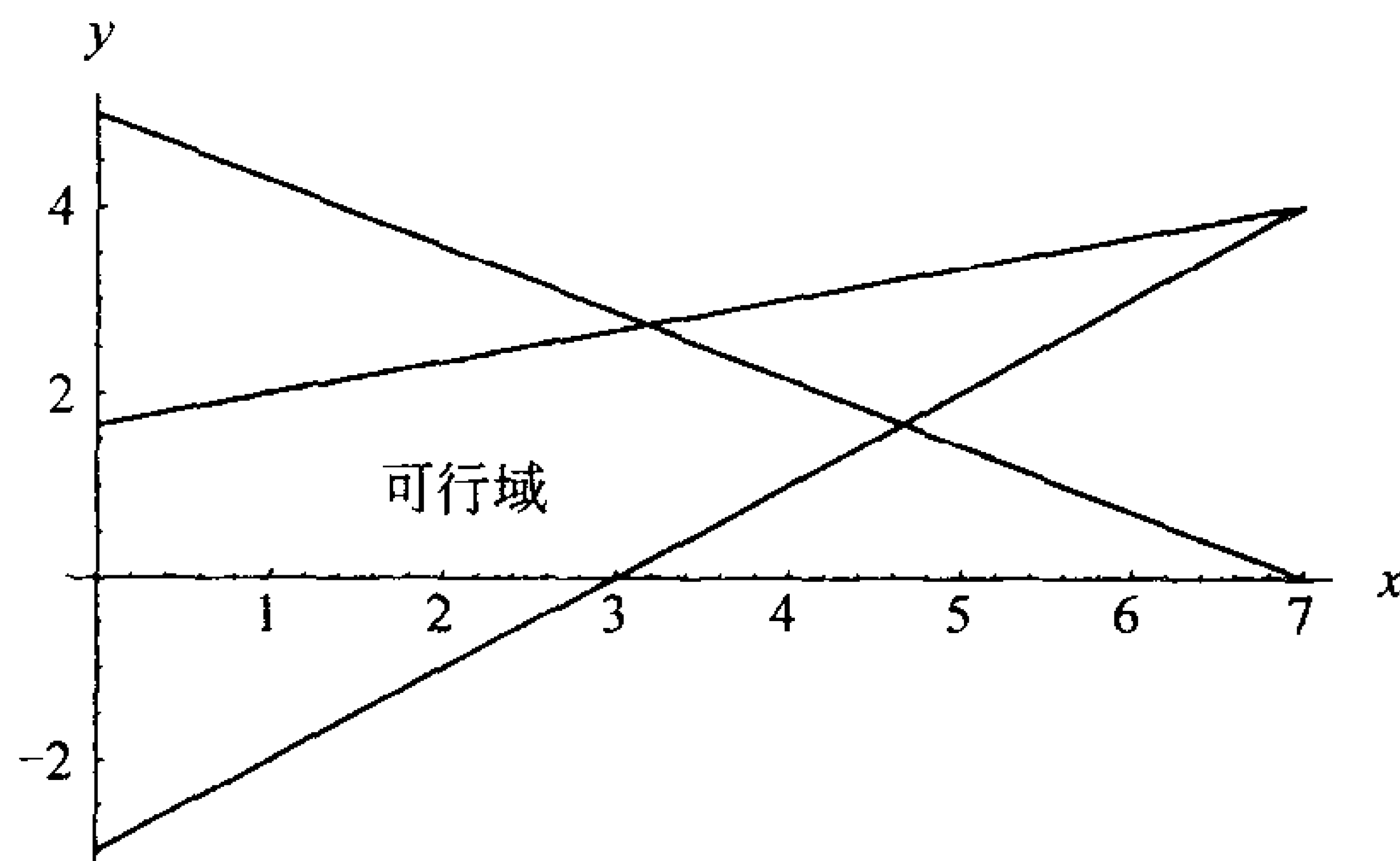


图3.7　可行域F

域F是一个凸（见第4章第12节）多边形，有5个顶点（一个5边形）．y坐

标等于 0 的顶点是(0，0)和(3，0)，y 轴上的顶点是(0，5/3)，第 4 个顶点是(14/3，5/3)，它相当于直线$x-y=3$ 和 $5x+7y=35$ 的交点，最高的顶点是直线 $x-3y=-5$ 和 $5x+7y=35$ 的交点，其坐标为$(x，y)=(35/11，30/11)$.

图 3.8 中在$(x，y)$平面上画出函数 $y=-x/9$ 的图形，它相当于目标函数的值 $f=0$.

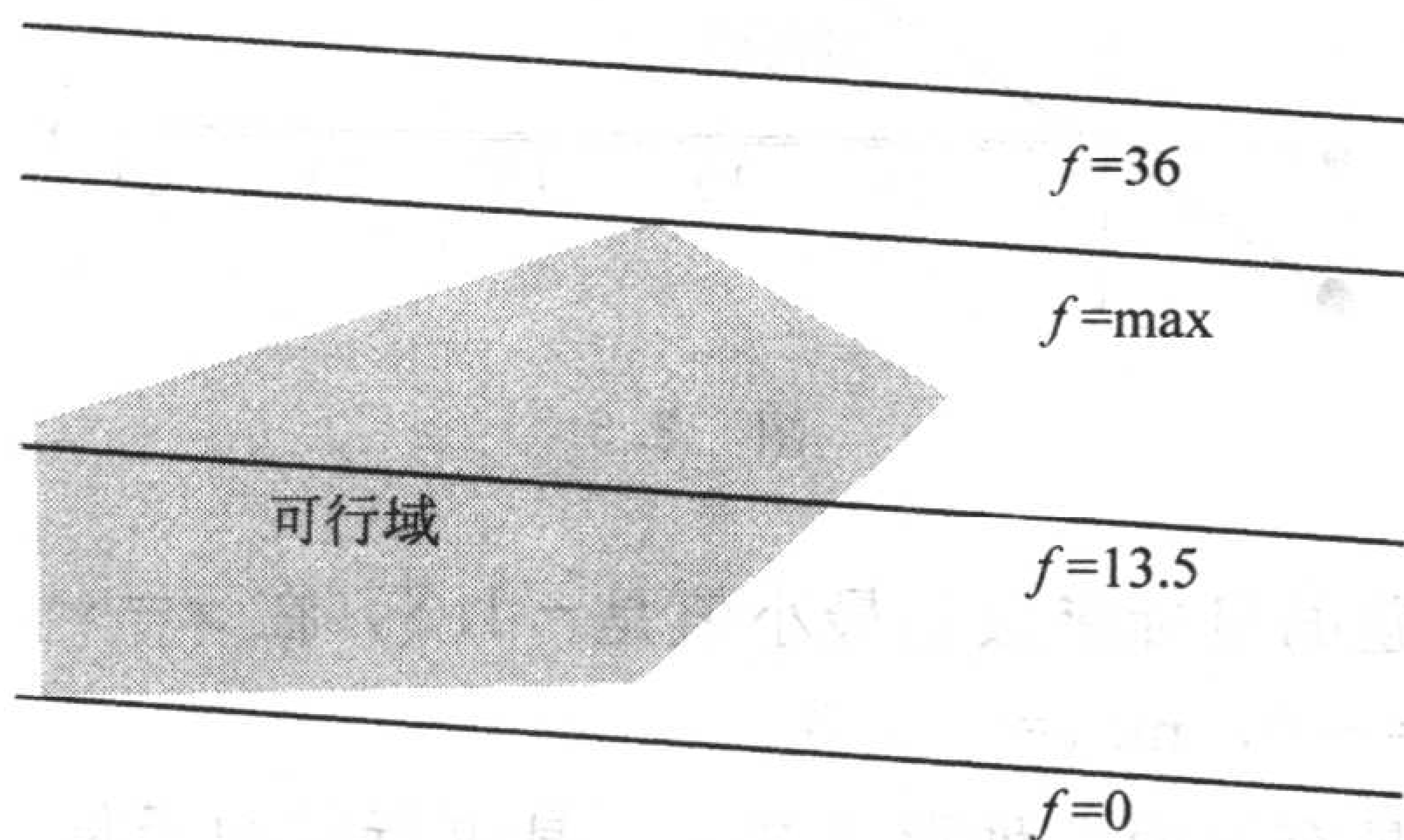


图 3.8　可行域 F 和 f 的等值线

28

沿着通过可行域的平行线 f 的值是常数，即 $x+9y=f$ $(f\geqslant0)$. 显然，f 的最小值是 0，$f=36$是不可行的，在 F 中 f 的最大值在顶点(35/11，30/11)达到，其最优值是 $f=35/11+9(30/11)=305/11$，目标函数 f 在 0 和 305/11 之间取值. 答案：$x=35/11$，$y=30/11$，$\max=305/11\approx27.7$. ■

一般地，有两个变量 x，y 的线性规划的可行域是下列形状之一的多边形区域：空集，一个点，一个区间，一条射线，一条直线，一个有界凸多边形，半平面，全平面，带域，角域，边数 $s\geqslant3$ 的无界多边形. 最优解集合是所列形状之一的子集.

当可行域是(有界、非空)多边形时，显然对于任何一个线性规划都至少有一个最优解，而且一个顶点(依赖于目标函数)是一个最优解(角点准则).

当然，角点准则对于非线性问题不成立. 下面是一个例子.

例 3.9　在例 3.6 的约束下求 $g=x^2-3x+2y^2-4y$ 的最小值.

在可行域内画出非线性目标函数的等值线(图 3.10).

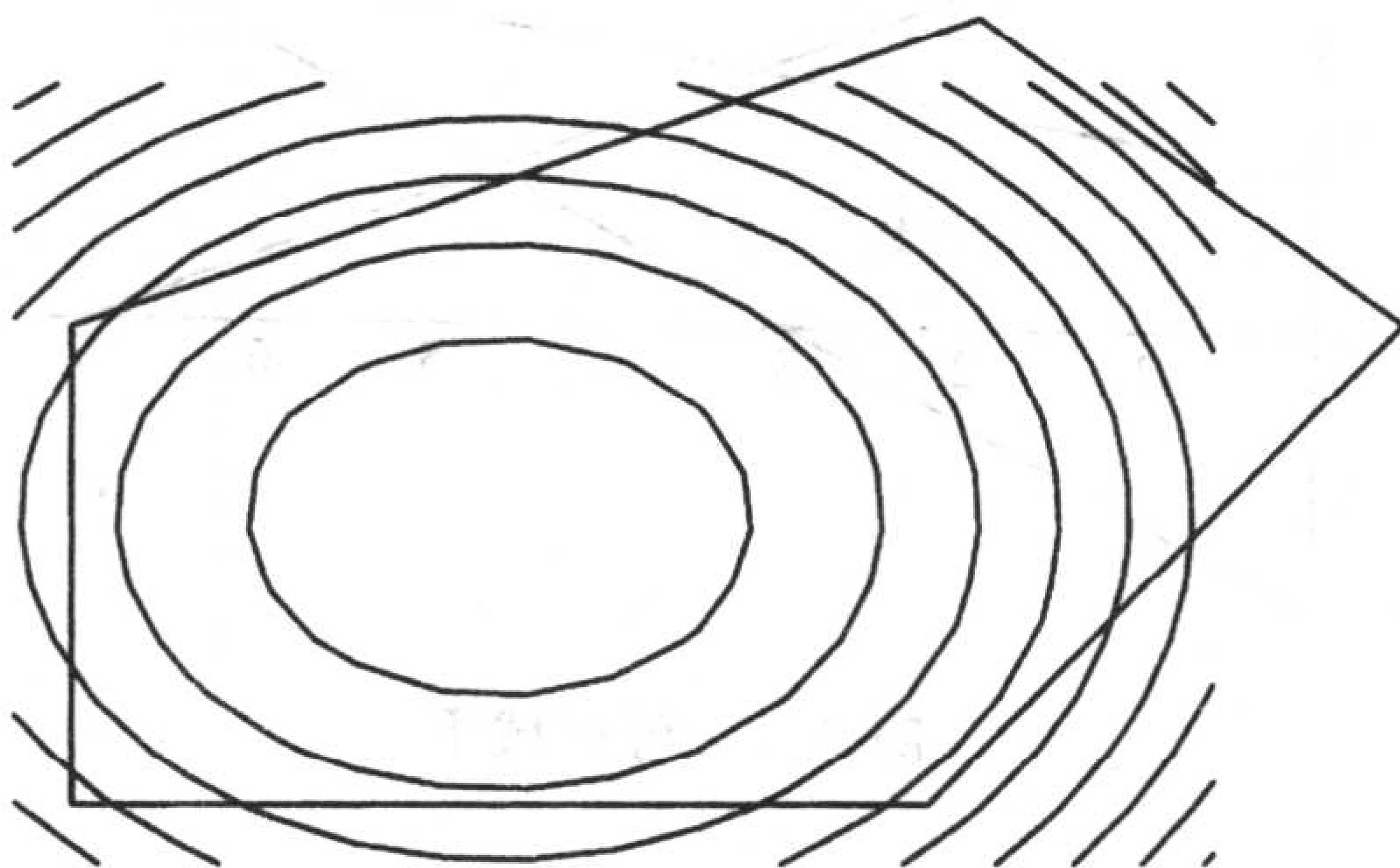


图 3.10　可行域 F 和例 3.9 中 g 的等值线

该图清楚地表明，最优解在 F 的内部，不在边界上，于是可以不管约束．为了在无约束下求 g 的最小值，可以利用导数或者作简单的代数变换：$g=(x-3/2)^2+2(y-1)^2-9/4-2$，因此在 $x=1.5$，$y=1$，min $=-4.25$． ■

[29]

从例 3.6 可知，一个非线性问题的另一种方式是增加非线性约束．

例 3.11　一个整数规划

对于例 3.6 的线性规划增加 x，y 都是整数的约束，然后求解．

再次作图(图 3.12)．

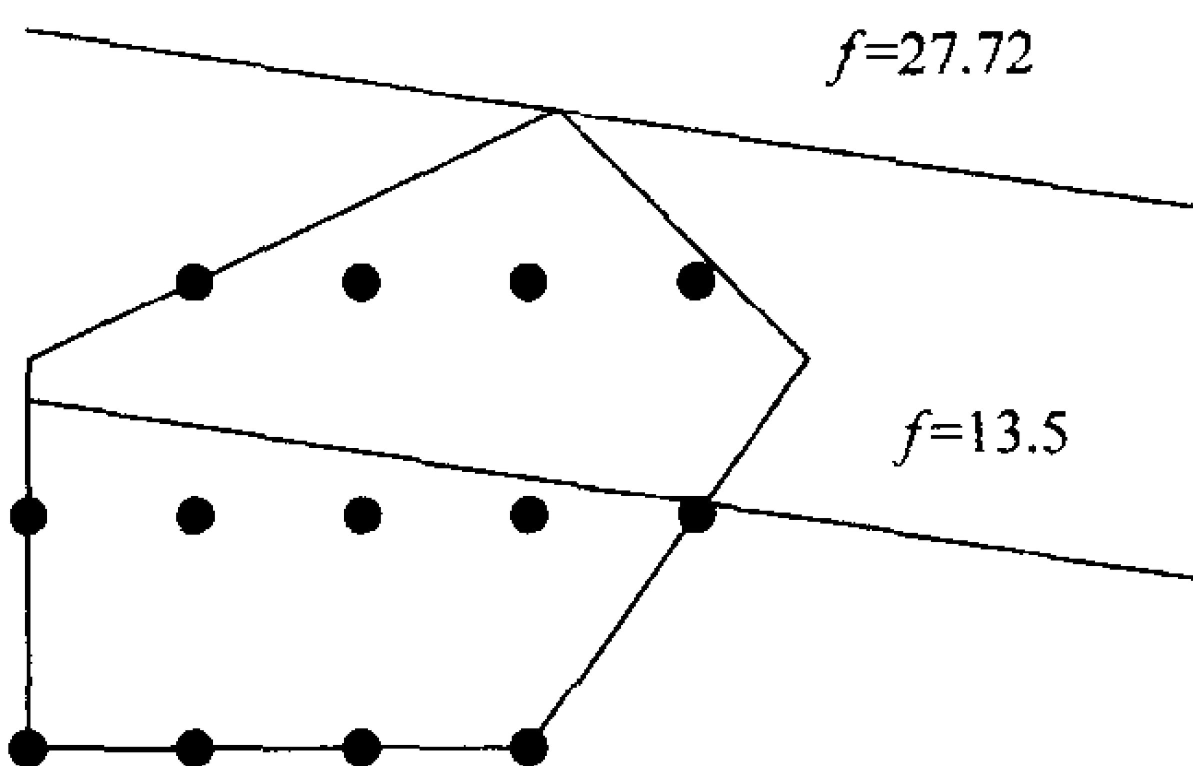


图 3.12　例 3.11 中的可行点和 f 的等值线

显然，可行域(F 中的整数点)由 13 个点组成，可能有两个点 $(x, y)=(1, 2)$，$(4, 1)$ 除外，它们显现在边界上．检查约束 $x-y\leqslant 3$(对 $x=4$，$y=1$)及 $x-3y\geqslant -5$(对 $x=1$，$y=2$)证实了图形的描述．从图 3.12 可以看出最优解是 $x=4$，$y=2$，max $=4+9\times 2=22$，增加的条件使最大值从 $305/11\approx 27.7$ 减少到 22．

一般来说，整数规划与线性规划有关，它的某些或全部变量是整数，这个附加的条件通常使问题的求解变得更加困难，但是本例还不是这样． ■

下面考虑例 3.6 的另一种改变——修改目标函数，增加非线性约束．

例 3.13

$$\begin{cases}\text{maximize} & h = x^2 - y^2\\ \text{subject to} & x\geqslant 0, y\geqslant 0, y\text{ 为整数},\\ & x-y\leqslant 3, x-3y\geqslant -5, 5x+7y\leqslant 35.\end{cases}$$

[30]

现在可行域由 3 条水平线段组成，非线性目标函数的等值线是分开的(图 3.14)．

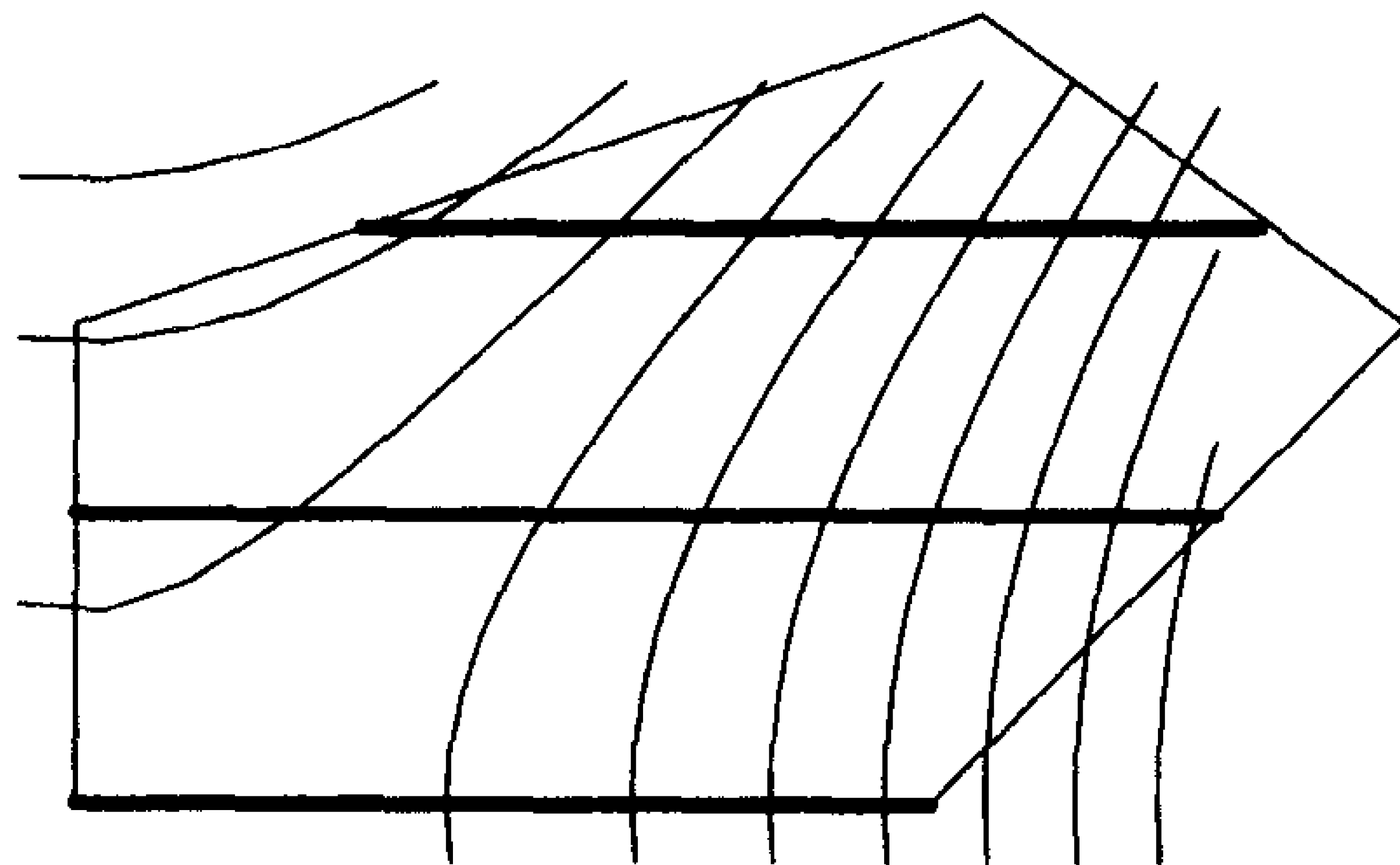


图 3.14　例 3.13 中的可行线和 h 的等值线

图形表明，最优解限于两个点：上面和中间区间的右端点，即 $x=4.25$，$y=2$ 和 $x=4$，$y=1$. 这两个点的 h 值是 14.062 5 和 15，于是在 $x=4$，$y=1$，max=15. ■

下面的例子是求解依赖于两个参数的一族线性规划.

例 3.15

maximize $x+y$ subject to $0\leqslant x\leqslant a$，$0\leqslant y\leqslant b$，其中 a，b 是给定的数.

显然这个问题的可行域是

一个矩形(若 a，$b>0$)；一条线段(若 $a=0$ 且 $b>0$，或 $b=0$ 且 $a>0$)；一个点(若 $a=b=0$)；空集(若 $a<0$ 或 $b<0$).

于是答案为

若 a，$b\geqslant 0$，则 $\max=a+b$，在 $x=a$，$y=b$ 达到. 其他情形下规划不可行.

这个问题可以不用图形求解，但是纸上或脑子里的图形是有帮助的. ■

对于变量个数很多的线性规划，可以把可行域和最优解集合想象成高维空间的凸集. 在定义了什么是凸集(见第 4 章第 12 节)之后，我们再将详细说明这个问题.

31

一个线性约束 $c_1x_1+\cdots+c_nx_n=a$ 给出一个*超平面*(hyperplane)(若不是所有的 $c_i=0$)，或者整个空间(若形如 $0=0$)，或者空集(若所有的 $c_i=0$ 但 $a\neq 0$). 线性约束 $c_1x_1+\cdots+c_nx_n\leqslant a$ 给出一个*半空间*(若不是所有的 $c_i=0$)，或整个空间(若所有 $c_i=0$，而 $a\geqslant 0$)，或空集(若所有 $c_i=0$，但 $a<0$).

借助计算机可以利用 3 维图形，它比纸有效得多. 用高维图形研究问题时纸不是好的媒介，而人们的想象则是.

练习

1～3. 试用手工完成这些练习，其中包含你的社会保险号码 $a_1a_2a_3a_4a_5a_6a_7a_8a_9$ 中的数字 a_i.

$$1.\begin{cases}\text{maximize} & (a_1-a_2)x\\ \text{subject to} & (a_3+a_4)x\leqslant a_5,\\ & (a_6+a_7)x\geqslant -a_8,\\ & (a_9+2)x\leqslant 10.\end{cases}$$

$$2.\begin{cases}\text{minimize} & f=a_1x-(a_1+a_2)y\\ \text{subject to} & |(9-a_3)x+a_4y|\leqslant 10+a_4,\\ & |a_5x+(1+a_6)y|\leqslant 8,\\ & |x+y|\leqslant a_7+a_9+1.\end{cases}$$

$$3.\begin{cases}\text{minimize} & f=a_1x+a_2y\\ \text{subject to} & |(9+a_3)x+a_4y|\leqslant 10,\\ & |a_5x+(9+a_6)y|\leqslant 10,\\ & |x+y|\leqslant a_7+a_8+a_9.\end{cases}$$

4～9. 求解下面的问题.

4. $$\begin{cases} \text{maximize(对 } x) & bx \\ \text{subject to} & |2x+4| \leqslant 10, \\ & |x+3| \leqslant 5. \end{cases}$$

其中：

(i) $b=7$.

(ii) $b=-9$.

(iii) b 是任意给定的数.

5. $$\begin{cases} \text{minimize} & f=x+8y \\ \text{subject to} & |x| \leqslant 9, \\ & |y| \leqslant 9, \\ & |x+y| \leqslant 9. \end{cases}$$

6. minimize x/y subject to $x \geqslant y$.

32

7. $$\begin{cases} \text{minimize} & f=x \cdot y \\ \text{subject to} & |x|+|y| \leqslant 1. \end{cases}$$

8. $$\begin{cases} \text{maximize} & w=x+y+z \\ \text{subject to} & |x| \leqslant 1, \\ & |y| \leqslant 1, \\ & |z| \leqslant 1, \\ & x+2y+3z=6. \end{cases}$$

9. maximize $\dfrac{1}{1+x^2+y^4}$.

10. 对线性方程 $bx=c$ 求解 b，其中 x，c 是给定的数.

11. $$\begin{cases} \text{maximize} & f=x+9y \\ \text{subject to} & x \geqslant 0,\ y \geqslant 0, \\ & x-y \leqslant 3, \\ & x-3y \geqslant -5, \\ & 5x+7y \leqslant 35, \\ & x,\ y \text{ 为整数}. \end{cases}$$

12. 求解食谱问题，其中包括 3 种成分——能量、维生素 B_1（硫胺）、维生素 B_2（核黄素），两种食物——杏仁（果仁，丢舍壳 60%）、蓝浆果（生的，丢舍 2%）. 下面是每 100g 可食用部分（不包括丢舍部分）的含量及价格（美元/lb 而不是美元/100g!）.

食物	能量(kcal)	B_1(mg)	B_2(mg)	价格
杏仁	578	0.24	0.81	2
蓝浆果	56	0.05	0.05	3
RDA	2 000	1.1	1.1	

13. maximize x^3+y^3 subject to $x+y\leqslant 5$.
14. maximize $|x|+y^2$ subject to $|x+y+2|+|x-y+3|\leqslant 5$.

33

15. maximize $x+2y+3z$ subject to $x,\ y,\ z\geqslant 0,\ x+y+z=1$.

第2章 背景知识

4. 逻辑

逻辑推理在线性规划中像在许多其他科学领域中一样重要. 只有学校里提给你的问题是现成的线性规划形式. 相反地，通常是先有数据，其中有些甚至是无关的，然后由你去综合与整理，直到形成要解的问题的类型. 回忆第 1 章第 2 节中的例子，开始时我们并不要求在一定的线性约束下求解一个线性型的最大值或最小值，而是要分析一个现实生活中的情况(虽然是简化的)，以便提出详细的行动计划. 根据分析的结果从中提炼出特定的线性规划问题需要逻辑推理.

为了逻辑推理必须了解某些名词的数学意义，它们可能与普通英文的习惯用法不同. 例如我们问，“你想喝咖啡或者茶?”我们用词“*或者*”(or)表示希望你从两种饮料中选择一种. 然而如果我们说“数 2 或者 3 是多项式方程 $x^2-x-6=0$ 的一个根”，是指这两个数都是根，数学上不是要从它们中进行选择. 换言之，在数学中“*或者*”通常表示一种可以相容的差别. 注意，在某些日常生活中也会有这种情况，例如有人问你，想在咖啡里加乳酪或者糖，你并非只能选择一种. 包含“*或者*”意义的一些数学符号是$\geq$，$\leq$，$\pm$.

[34] 有着比英文中更精确的数学定义的另一个看来简单的词是“*和*”(and). 考虑这个词的各种使用方式. 当你列出一系列对象时用到它：“我买了一条裙子、一件羊毛衫和一件外衣.”还有“我洗了头和刷了牙.”然而数学意义下的“*和*”，是指两种(或更多)陈述必须同时满足. 例如“*和*”在以下语句中的应用：

“b 是一个正整数和方程 $x^2-x-6=0$ 的一个根”

表明 b 一定等于 3. 为什么?

如果在数学意义上熟练地应用下面这些词，并熟悉它们的逻辑符号，将会提高数学理解能力：

- 和(逻辑符号：$\wedge$).
- 或(逻辑符号：$\vee$).
- 蕴涵(逻辑符号：$\Rightarrow$).
- 来自(逻辑符号：$\Leftarrow$).

有时，符号$\Rightarrow$和$\Leftarrow$组合成一个符号$\Leftrightarrow$，它与“当且仅当”、“即”、“等价于”、“意思是”、“$\Rightarrow\wedge\Leftarrow$”的意义相同.

除了逻辑等价之外，在数学及其他地方还有另外一些等价关系. 在说“这些定理是等价的”时，常常是对“这两个定理或者都对，或者都错”的不同形式表述.

通常，逗号表示“和”. 例如，$x>0$，$y>0$ 表示 $x>0$ 和 $y>0$，也可写作 $x, y>0$. 线性规划中的约束常用逗号连接，表示它们都要满足. 但是某些情形下逗

号的意思是另一回事，如$x=0$，1，或2表示x取这3个值中的一个. 在例子$x=2+3$，或$x=5$中逗号把“或”的意思改成“即”(i. e.).

通常有很多词在数学上有相同的意义，如在真实的陈述

35

$$x \geqslant 0 \text{ 如果 } x = 1 \tag{4.1}$$

中，可以用下面任何一个词或符号代替*如果*：

当，由于，只要，弱于，来自，是……的结果，因为，被蕴涵于，$\Leftarrow$[㊀].

特别地，(4.1)可以写作：$x\geqslant 0 \Leftarrow x=1$，也可写作：

$$\text{如果 } x = 1\text{，则 } x \geqslant 0 \text{，}$$

或者等价地

$$x = 1 \text{ 蕴涵 } x \geqslant 0. \tag{4.2}$$

可以用下面任何一种表达方式代替(4.2)中的*蕴涵*：

强于，只有当，于是，导致，给出，所以，因此，$\Rightarrow$[㊁].

例 4.3 下面14个句子的数学意义相同：

- $x\geqslant 0$；因为$x\geqslant 2$.
- $x\geqslant 0$，如果$x\geqslant 2$.
- $x\geqslant 2$仅当$x\geqslant 0$.
- 如果$x\geqslant 2$，则$x\geqslant 0$.
- 界限$x\geqslant 2$比$x\geqslant 0$更严.
- 已知$x\geqslant 2$，则$x\geqslant 0$.
- 界限$x\geqslant 2$比$x\geqslant 0$更好.
- 约束$x\geqslant 0$比$x\geqslant 2$更松.
- 界限$x\geqslant 2$比$x\geqslant 0$更精确.
- 由于$x\geqslant 2$，所以$x\geqslant 0$.
- $x\geqslant 2$蕴涵$x\geqslant 0$.
- 约束$x\geqslant 0$比$x\geqslant 2$更弱.
- $x\geqslant 2$是$x\geqslant 0$的充分条件.
- 约束$x\geqslant 0$来自条件$x\geqslant 2$. ■

注 *在定义中，有时“如果”等价于“当且仅当”.*

本书中许多陈述或条件是对变量的约束，每一个约束或约束组给出一个集

㊀ 原文给出的词汇是：when，since，provided that，whenever，is weaker than，follows from，is a consequence of，because，is implied by，$\Leftarrow$. 这里没有一一译出. ——译者注

㊁ 原文给出的词汇是：is stronger than，only if，so，hence，results in，forces，gives，whence，therefore，thus，consequently，$\Rightarrow$. 这里没有一一译出. ——译者注

合——可行域．对应于较强约束组的可行域是对应于较弱约束组的可行域的一部分，增加新的约束会缩小可行域．例如，点 $x=1$ 属于射线 $x\geqslant 0$，表示条件 $x=1$ 蕴涵 $x\geqslant 0$．等价的约束组彼此蕴涵，它们有相同的可行域．对于可行域"和"表示交集，"或"表示并集，例如条件"$x=0$ 和 $y=0$"给出平面上一个点，而条件"$x=0$ 或 $y=0$"给出两条直线的并． 36

下面是一个更复杂的例子．考虑这样的陈述：$0=1$ 蕴涵 $0=0$．这个陈述是对的，因为任何错误陈述通常蕴涵每件事情．同样，正确陈述通常来自每件事情．特别地，我们不需要任何条件就可得到 $0=0$．说明上面是正确陈述的另一种办法是下面的事实：给定 $0=1$，可以得到 $0=1=0$，故 $0=0$．用 0 去乘等式 $0=1$ 也得到所要的结果．最后，上面的陈述用可行集合的术语可以表示为：空集是全集的一部分．

现在考虑更复杂的线性方程．给出两个线性方程

$$\begin{cases} x+3y=3 \\ x+2y=5, \end{cases}$$

可以取它们的差得到 $y=-2$，于是方程 $y=-2$ 来自这个方程组．更一般地，给定 m 个方程 $f_1=b_1$，…，$f_m=b_m$，及任意的数 c_i，取所给方程与系数 c_i 的线性组合：

$$c_1 f_1+\cdots+c_m f_m=c_1 b_1+\cdots+c_m b_m.$$

这个方程逻辑上来自上面的方程组，它表明方程组的每个解满足该方程．而一旦知道了如何解线性方程组，就容易说明其逆陈述是对的，只要方程组是相容的，并且所有方程都是线性方程．

也就是说，给定线性方程组 $f_1=b_1$，…，$f_m=b_m$，以及来自这个方程组并以标准形式表示的另一个方程 $f_0=b_0$，则或者方程 $f_0=b_0$，或者方程 $0=1$ 是该方程组中方程的一个线性组合．这个问题的一般情形见本章第 6 节，特殊情形见练习 38～41．对于线性不等式而言类似陈述也是对的，只要仔细地对待系数的符号（见第 5 章第 15 节）．

让我们回到关于逻辑的一般性讨论，没有办法把逻辑符号的所有等价表示以及它们的用法和误用全列出来，下面是一些常见的谬论： 37

- $x\geqslant 0\Rightarrow x=0$ 或 1（错误的二难推理）．
- $0=0$ 于是 $0=1$（无知的论据）．
- $x=1/3$ 或作为十进位小数 $x=0.333\,33$（滑坡效应）．
- $1=1$ 或 $0=1$ 是错的（复杂的结论）．
- $0>1$ 因为这是我的教师教的（请求外援）．
- $0>1$ 因为这是我通宵达旦得到的结果（请求同情）．
- 方程 $x^2=-1$ 中的负号是一个错误，否则解它只会浪费时间（要求推理）．

- 每个讲道理的人都同意 1/3=0.3(有偏见的语言).
- 每个人都知道 1/3=0.3，因此它是正确的(请求普及性).
- 我不能同意 0=1 是一个线性约束，因为你甚至不能写出约束(攻击此人).
- 线性方程有一个实数解，因为《纽约时报》这样写并且专家同意(请求权威和匿名专家).
- 第 1 章中所有例子的约束都少于 10 个，于是每个线性规划的约束都少于 10 个(草率的推广).
- 第一周课我没有困难，因此这门课是一块蛋糕，我不需要努力就可以通过(没有代表性的例证).
- 为了解线性方程组我对方程作加减法，于是我用同样的运算简化线性约束系统(错误的模仿，虽然方程可以求和，但是对于不同类型的不等式却不行).
- 每个含一个未知数的线性方程有一个解(排除的假象).
- 这一节在第 1 章之后，因此我们在第 1 章中不需要任何逻辑，但是需要知道什么样的线性规划是合乎逻辑的(巧合的关联).
- 由于在我的优化问题中所有约束都是线性的，其目标函数必定是仿射的(连锁影响：对线性规划而言假设和结论二者都对).
- $x>0$ 导出 $x>1$(方向错误).
- 复数不是真正的数(模棱两可的表达).
- 方程 0=1 没有解(模棱两可，即两个不同的意思).
- 线性规划中数据和解都是实数，所以每个线性规划都有一个可行解(存在的谬误).

38

如果你的头脑还没有眩晕，你认为这个陈述中的逻辑是什么："由于我确实理解了包含 10 类谷物的食谱问题，从现在起我只吃一种谷物."更多的例子请看 http://www.intrepidsoftware.com/fallacy/toc.htm.

学习逻辑就像学习走路，需要耐心和实践．用一些引语作为本节的结束，让读者去判断其中是否有任何逻辑.

Jacques Hadamard (1865—1963)：

逻辑只认可直觉的战利品.

Antoine Arnauld (1612—1694)：

普通的意义实际上并不如此普通.

Ludwig Wittgenstein (1889—1951)：

逻辑中决不会有惊奇.

Morris Kline（生于 1908 年）：

逻辑是满怀信心地犯错误的艺术.

Lord Dunsany（1878—1957）：

逻辑像威士忌一样，接受过多就会失去它的有益作用.

Oliver Heaviside（1850—1925）：

逻辑是宽容的，因为它是永恒的.

怎么可以只因为不了解相关的消化过程，就拒绝一顿丰盛的晚宴呢?

G. K. Chesterton（1874—1936）：

如果没用逻辑已经发现了真理，那么用逻辑也只能发现真理.

Hermann Weyl（1885—1955）：

逻辑是数学家锻炼思维保持其健康和强壮的保健法.

Richard Feynman（1918—1988）：

……数学不只是一种语言……它是一种语言加逻辑. 数学是推理的一种工具.

Jeremy Bentham(1748—1832)：

啊，逻辑：你生来就是科学殿堂的守门人，却成了反复无常的命运的牺牲品，至今难以避免替学究们做苦工的厄运. 赶快请求你主人的帮助吧!

39

Jules Henri Poincaré(1854—1912)：

靠逻辑证明，靠直觉创造. 这样，为了证明一个已经理解的定理，弄懂它有什么意思，既没有必要，也没有好处. 几何学家可以用 Stanley Jevons 想象的"逻辑钢琴"(logic piano)代替；或者，如果喜欢，也可以想象成一台机器. 假设从机器的一头进去，定理从另一头出来，就像传说中的 Chicago 机器，生猪活着进去，变成火腿和香肠出来. 数学家应该知道他所做的事不比这些机器更多.

Bertrand Russell(1872—1970)：

日常用语一般来说不适合表达物理学的真正主张，因为日常用语不够抽象. 只有数学和数学逻辑能够说得像物理学家想要说的一样少.

David van Dantzig(1900—1959)：

无论在主观世界还是客观世界里我们都不能找到判断数的概念的真实性标准，因为主观世界里不包含这样的概念，而客观世界里除了受这个概念影响的事物以外就什么也没有了. 那么我们怎样得到一个标准呢? 不能靠证据，因为证据的骰子的一边已经被加重了. 不能靠逻辑，因为逻辑不能离开数学独立存在，数学只是增加其必要性的一个方面. 对数学

的概念又如何判断呢？根本就无法判断. 数学是至高无上的仲裁者，对于数学的结果不存在上诉. 我们不能改变这个游戏规则，也不能确定它是否公平，我们只能研究这个游戏中的比赛者，然而不是以旁观者的态度，因为我们需要关注自己在游戏中的意向.

练习

1～31. 确定下面 30 个陈述的正确性，如果同意记“对”，否则记“错”，解释原因.

1. $|x|=1$ 仅当 $x\geqslant 0$.
2. $xy=0$ 仅当 $y=0$.
3. $|x|\leqslant 1$ 如果 $x\leqslant 1$.
4. 如果 $|x|\leqslant 1$，则 $x\geqslant -1$.
5. $x\neq 1$ 除非 $x\geqslant 0$.
6. $|x|>1$ 因此 $x>0$.
7. $x\geqslant 0$ 只要 $x>2$.
8. 如果 $0=1$，则 $2=5$.
9. $x\geqslant 0 \Leftarrow x>2$.

40

10. $x^2=0 \Leftrightarrow x=0$.
11. $x=1$ 不蕴涵 $x\geqslant 1$.
12. $|x|=1$ 强于 $x\geqslant 0$.
13. $x\geqslant 0$ 来自 $x=5$.
14. $|x|\leqslant 1$ 当且仅当 $x\leqslant 1$ 且 $x\geqslant -1$.
15. $|x|\leqslant 1$ 当且仅当 $x\leqslant 1$ 或 $x\geqslant -1$.
16. $x^2=0$ 等价于 $x=0$.
17. $x=y$ 等价于 $x-y=0$.
18. $|x|\geqslant 1$ 则或者 $x\leqslant -1$ 或者 $x\geqslant 1$.
19. $|x|\geqslant 1$ 当且仅当 $x\leqslant -1$ 和 $x\geqslant 1$.
20. 如果 $xy=0$ 则 $x=0$ 或 $y=0$.
21. $x\geqslant 1$，$x\geqslant 0$ 中的第 2 个条件是多余的.
22. 给定 $0=1$，可得 $2=5$.
23. $x>10$ 是 $x\geqslant 0$ 的充分条件.
24. $x=3$ 导出 $x\geqslant 0$.
25. $x^2>10$ 则 x 为正.
26. 如果 $|x-1|>4$，则 $x>5$ 或者 $x<-3$.
27. 如果 $x>y$，则 $-x<-y$.
28. $x>0$ 是 $x>2$ 的必要条件但不充分.

29. 条件 $x=1$ 比条件 $x\geqslant 1$ 弱.

30. 条件 $x=1$ 可由 $x\geqslant 1$ 给出.

31. $|x|\leqslant 1$ 当且仅当 $x\leqslant 1$ 且 $x\geqslant -1$.[⊖]

32～39. 从下面 4 个条件中找出所有蕴涵的条件：

32.

(i) $x=2$，$y=3$

(ii) $x\geqslant 0$

(iii) $y\geqslant 0$

(iv) $x+y=5$

33.

(i) $x=2$，$y=3$ 或 $x=y=0$

(ii) x，$y\geqslant 0$；

(iii) x，y 为整数

(iv) $|x+y|\leqslant 5$

34.

(i) $0=1$

(ii) $0=0$

(iii) $1=2$

(iv) $x=y$

35.

(i) $x=1$，$y=3$

(ii) $x+y=3$，$x-y=1$

(iii) $x=1$ 或 $y=2$

(iv) $2x+3y=8$

36.

(i) $x+y\geqslant 1$，$x-y\geqslant 2$

(ii) $2x\geqslant 3$

(iii) $2y\geqslant -1$

(iv) x，$y\geqslant 0$

37.

(i) $x=0$

(ii) $x^3=0$

(iii) $0=0$

(iv) $xy=0$

⊖ 与第 14 题重复. ——译者注

38.

(i) $x^2+y^2=1$

(ii) $x^2+y^2\leqslant 1$

(iii) $x\leqslant 1$，$y\leqslant 1$

(iv) $x+y\leqslant 2$

39.

(i) $|x|+|y|\leqslant 1$

(ii) $x^2+y^2\leqslant 1$

(iii) $|x|\leqslant 1$，$|y|\leqslant 1$

(iv) $x^4+y^4\leqslant 1$

40. 对于式(4.1)中的“如果”，有哪些其他可能的替代词?

41. 对于式(4.2)中的“蕴涵”，有哪些其他可能的替代词?

42～48. 下列陈述正确吗?

42. 这个优化问题是线性的，所以它的目标函数必定是线性的.

43. 如果在线性规划中每一个约束是可行的，则这个规划是可行的.

44. 如果一个线性问题是不可行的，则所给约束之一必定是不可行的.

45. 如果 $x=3$ 且 $y=-1$，则 x 比 y 更接近 0.

46. $x\geqslant 0$ 因为 $x>y$ 且 $y>3$.

47. 约束 $a+2b+3c\geqslant 2$ 来自不等式组

$$a+b+c\geqslant 1, b+2c\geqslant 1.$$

48. 约束 $a+2b+3c\leqslant 3$ 来自不等式组

$$a+b+c\leqslant 1, b+2c\leqslant 1.$$

49～52. 检查第 3 个方程是否来自前两个方程(即它在该方程组中是否多余)，如果是，将它写成前两个方程的线性组合.

49. $\begin{cases} x=0, \\ 2y=5, \\ x+y=2. \end{cases}$

50. $\begin{cases} x+y+z=1, \\ x+2y+3z=3, \\ 2x+3y+4z=4. \end{cases}$

51. $\begin{cases} a-b+c+d=0, \\ 2a+b-3c=1, \\ 3b-5c-2d=1. \end{cases}$

52. $\begin{cases} a-b+c+d=0, \\ 2a+b-3c=1, \\ -a-2b+4c+d=-1. \end{cases}$

5. 矩阵

线性代数和线性规划中经常用到矩阵，它使人们能以一种简化的形式书写线性方程组和不等式组．我们将从矩阵的定义开始，然后介绍如何借助矩阵用一种简单的记号书写线性规划问题．

定义 5.1 矩阵(matrix)是元(entry)的一个矩形排列，这些元可以是数、变量、多项式、函数等．

一般来说，一个矩阵 A 可以写成以下形式：

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

其中，$[a_{11} \quad a_{12} \quad \cdots \quad a_{1n}]$称为矩阵的第 1 行，$[a_{21} \quad a_{22} \quad \cdots \quad a_{2n}]$是矩阵的第 2 行，依次类推．类似地，

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

是第 1 列，

$$\begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}$$

是第 2 列，依次类推．此时我们说矩阵 A 有 m 行和 n 列，A 称为 $m\times n$ 矩阵．用 a_{ij} 记位于第 i 行、第 j 列的元．

当 $m=1$ 时(即矩阵只有一行)A 是行矩阵，或者就是行，也称行向量，或者就是向量．当 $n=1$ 时 A 是列矩阵，或者就是列，也称列向量，或者就是向量．

> **注** a_{ij} 的下标表示一对数，而不是乘积．当 m，n 较大时用 $a_{12,3}$ 或 $a_{1,23}$ 代替 a_{123}．在行矩阵中常用逗号来避免混淆，试比较[1 2 34]和[1, 2, 34]．

例 5.2

(i) $A=\begin{bmatrix} 1 & -1 & 3 \\ \sqrt{2} & \pi & 6 \\ -4 & 56 & 20 \end{bmatrix}$是实元构成的 3×3 矩阵．

(ii) $B=\begin{bmatrix}1 & x & x^2\\ 1 & y^2 & y^4\end{bmatrix}$是多项式元构成的 2×3 矩阵.

(iii) $\begin{bmatrix}x\\ y^2\end{bmatrix}$是 B 的第 2 列.

(iv) $C=\begin{bmatrix}\sin x & e^x & \ln x & x+1\\ \sqrt{x^2-\pi^2} & 3/4 & 1/x & 5\\ e^{7\cos x} & \tan\pi x & 2 & 5\end{bmatrix}$是函数元构成的 3×4 矩阵.

(v) $\begin{bmatrix}\sqrt{x^2-\pi^2} & \frac{3}{4} & \frac{1}{x} & 5\end{bmatrix}$是 C 的第 2 行.

(vi)若记 c_{ij} 为 C 的元，其中 $1\leqslant i\leqslant 3$，$1\leqslant j\leqslant 4$，则因为 $\tan\pi x$ 在第 3 行第 2 列，所以它是 c_{23}. ■

定义 5.3 两个矩阵 A 和 B 相等，如果它们大小相同[⊖]，并且 A 的每个元等于 B 的相应的元.

在线性规划中，同样大小的行(或列)[⊜]之间也用不等式. 除非另外指出，否则对于两个同样大小的向量，$A\geqslant B$ 表示 A 的每个元大于或等于 B 的相应的元.

回顾有理数集或实数集的引入，在这些数集定义之后，加减乘除运算及其性质即可阐明. 例如，加法和乘法都是可交换和可结合的，通过称之为乘法对加法的可分配性可以将加法和乘法连接起来. 让我们对新定义的称为矩阵的对象做类似的事情.

44

定义 5.4 两个 $m\times n$ 矩阵 A 和 B 的加法运算以分量形式定义，即 $m\times n$ 矩阵 $A+B$ 的第 i 行、第 j 列元是 $a_{ij}+b_{ij}$. 减法类似地定义.

容易验证，矩阵加法是可交换和可结合的.

定义 5.5 两个矩阵 A 和 B 的乘积仅当 A 的列数等于 B 的行数时才可定义，若 A 是 $m\times n$ 矩阵，B 是 $n\times p$ 矩阵，则乘积 $A\cdot B$(或简记为 AB)是 $m\times p$ 矩阵，它的第 i 行第 j 列元是

$$(A\cdot B)_{ij}=\sum_{k=1}^{n}a_{ik}b_{kj}.$$

由两个矩阵乘积定义的法则可知，这个运算是不可交换的. 相反顺序的乘积 $BA=B\cdot A$ 可能无法定义，即使能够定义，也可能与 $A\cdot B$ 不同.

然而矩阵乘法是可结合的. 什么意思呢? 设 A 是 $m\times n$ 矩阵，B 是 $n\times p$ 矩阵，C 是 $p\times q$ 矩阵，则乘积 $A\cdot B$，$(A\cdot B)\cdot C$，$B\cdot C$ 及 $A\cdot(B\cdot C)$ 均有定

⊖ 指行数 m 和列数 n 都相同. ——译者注

⊜ 指行(或列)的元的数目相同. ——译者注

义，且

$$(A \cdot B) \cdot C = A \cdot (B \cdot C).$$

回顾关于实数性质的知识，乘法对加法的分配性给出等式

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

矩阵乘法对加法的分配性也成立，只要适当的矩阵之和与矩阵之积有定义. 例如，若 A 是$m \times n$矩阵，B 和 C 是两个 $n \times p$ 矩阵，则 $A \cdot (B + C)$ 与 $A \cdot B + A \cdot C$ 得到相同的 $m \times p$ 矩阵，因此

$$A \cdot (B + C) = A \cdot B + A \cdot C.$$

这些结论的证明可以在任何的线性代数教科书中找到，这里从略. 下面是更多的例子.

45

例 5.6 设 A 和 B 是下面的矩阵：

$$A = \begin{bmatrix} 1 & -1 & 3 \\ 2 & 1/2 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1/2 & -3/4 & 7 \\ 2 & 3 & 8 \end{bmatrix},$$

则它们的和 $A+B$ 与它们的差 $A-B$ 是

$$A + B = \begin{bmatrix} 3/2 & -7/4 & 10 \\ 4 & 7/2 & 14 \end{bmatrix}, \quad A - B = \begin{bmatrix} 1/2 & -1/4 & -4 \\ 0 & -5/2 & -2 \end{bmatrix}.$$

■

例 5.7 设 C 和 D 是下面的 2×3 和 3×3 矩阵：

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad D = \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix},$$

则乘积 $C \cdot D$ 是下面的 2×3 矩阵：

$$\begin{bmatrix} 1.9+2.6+3.3 & 1.8+2.5+3.2 & 1.7+2.4+3.1 \\ 4.9+5.6+6.3 & 4.8+5.5+6.2 & 4.7+5.4+6.1 \end{bmatrix}$$

$$= \begin{bmatrix} 30 & 24 & 18 \\ 84 & 69 & 54 \end{bmatrix}$$

■

注意，乘积 $D \cdot C$ 没有定义，为什么？

例 5.8 设 A 是 2×3 矩阵

$$A = \begin{bmatrix} 1 & -2 & 0 \\ -3 & 4 & 1 \end{bmatrix},$$

x 是 3×1 列矩阵

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

且 b 是 2×1 列矩阵

$$b=\begin{bmatrix}-5\\6\end{bmatrix},$$

则乘积 $A\cdot x$ 是有定义的 2×1 列矩阵，故矩阵等式 $A\cdot x=b$ 有意义，并且在进

46

行这些运算后我们得到线性方程组

$$\begin{cases}x_1-2x_2=-5\\-3x_1+4x_2+x_3=6.\end{cases}$$

反之，这个线性方程组可以写成矩阵形式 $A\cdot x=b$. 综上所述可得下面基本而又重要的结果：

任何线性方程组可以写成标准的矩阵形式

$$Ax=b,$$

其中 x 是未知数(变量)构成的列，A 是给定的矩阵(系数矩阵)，b 是已知数的列(常数项). ■

注意，方程的个数不一定等于未知数的个数. 换言之，系数矩阵不一定是方阵.

定义 5.9 $m\times n$ 矩阵 A 的转置矩阵是 $n\times m$ 矩阵，它的(i, j)位置的元是原矩阵 A 的(j, i)位置的元. 记 A 的转置为 A^{T}，于是 $(A^{\mathrm{T}})_{ij}=A_{ji}$.

命题 设 A 是 $m\times n$ 矩阵，B 是 $n\times p$ 矩阵，可以定义 $A\cdot B$ 和 $B^{\mathrm{T}}\cdot A^{\mathrm{T}}$，且$(A\cdot B)^{\mathrm{T}}=B^{\mathrm{T}}\cdot A^{\mathrm{T}}$.

证明概要

(i) 计算乘积 $A\cdot B$.

(ii) 求矩阵 $A\cdot B$ 的转置.

(iii) 计算乘积 $B^{\mathrm{T}}\cdot A^{\mathrm{T}}$.

(iv) 比较矩阵$(A\cdot B)^{\mathrm{T}}$ 和 $B^{\mathrm{T}}\cdot A^{\mathrm{T}}$ 的相应元. ■

例 5.10 3×4 矩阵

$$C=\begin{bmatrix}\sin x & \mathrm{e}^x & \ln x & x+1\\ \sqrt{x^2-\pi^2} & 3/4 & 1/x & 5\\ \mathrm{e}^{7\cos x} & \tan\pi x & 2 & 5\end{bmatrix}$$

的转置是下面的 4×3 矩阵：

$$C^{\mathrm{T}}=\begin{bmatrix}\sin x & \sqrt{x^2-\pi^2} & \mathrm{e}^{7\cos x}\\ \mathrm{e}^x & 3/4 & \tan\pi x\\ \ln x & 1/x & 2\\ x+1 & 5 & 5\end{bmatrix}.$$

47

■

实数 0 和 1 分别是加法和乘法的特殊元，因为 0 加一个数不影响这个数，类似地，用 1 乘一个数得到原来的数. 我们定义有类似性质的矩阵.

定义 5.11 零矩阵是只含零元的 $m\times n$ 矩阵.

注 零矩阵的大小通常从其所处环境可以清楚地知道，不需要特别指出. 如不等式$[0, 1, 2, 3]\geqslant 0$ 的第 1 个 0 表示一个数，而第 2 个 0 可以理解为 1×4 矩阵. 在不等式

$$[1, -1, 2]\begin{bmatrix}3\\5\\1\end{bmatrix}=0$$

中右边的 0 表示 1×1 矩阵，即一个数.

$m\times n$ 矩阵 A 的加性逆(additive inverse)是由 A 的元的加性逆构成的 $m\times n$ 矩阵. 记 A 的加性逆为 $-A$，于是对于任意大小的矩阵 A 都有 $A+(-A)=0$. 相反地，不是每个矩阵都有乘法逆，回顾即使对于实数，0 并不可逆.

定义 5.12 对正整数 n，定义 $n\times n$ 单位矩阵 I_n 为对角元是 1、其他元是 0 的矩阵.

矩阵$[a_{ij}]$的对角元是 a_{ii}，单位矩阵 I_n 是正方对角矩阵. 一般来说，对角矩阵定义为所有非对角元均为 0 的矩阵.

下面是 2×3 对角矩阵的例子：

$$\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\end{bmatrix}.$$

下面是两个非对角矩阵(为什么?)的例子：

$$\begin{bmatrix}1 & 2\\0 & 1\\5 & 0\end{bmatrix},\quad \begin{bmatrix}0 & 1 & -1 & \pi & 1/2\\0 & 0 & 1 & 3 & 4\\0 & 0 & 0 & 0 & 1\\0 & 0 & 0 & 0 & 0\end{bmatrix}.$$

48

下面是 2×2 单位矩阵 I_2 的例子：

$$I_2=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}.$$

下面是 5×5 单位矩阵的例子：

$$I_5=\begin{bmatrix}1 & 0 & 0 & 0 & 0\\0 & 1 & 0 & 0 & 0\\0 & 0 & 1 & 0 & 0\\0 & 0 & 0 & 1 & 0\\0 & 0 & 0 & 0 & 1\end{bmatrix}.$$

定义 5.13 $n\times n$ 矩阵 A 称为可逆(invertible)的，如果存在 $n\times n$ 矩阵 B 使 $A\cdot B=B\cdot A=I_n$.

命题 5.14 设 A 是可逆的 $n\times n$ 矩阵，则 A 的逆是唯一的.

证明 设对可逆矩阵 A 有两个逆矩阵 B 和 C，按照定义 $A\cdot B=I_n$，$C\cdot A=I_n$. 因此由 $C=C\cdot I_n$ 和 $A\cdot B=I_n$，得到 $C=C\cdot(A\cdot B)$. 由于矩阵乘法是可结合的，上式的乘积项等于 $(C\cdot A)\cdot B$，因 $C\cdot A=I_n$ 它又等于 $I_n\cdot B=B$. 联合以上的等式，得到 $C=B$. ■

从现在起用记号 A^{-1} 表示 $n\times n$ 可逆矩阵 A 的逆矩阵，在用矩阵的这种表示之前有几个需要考虑的问题：

(i)如何计算一个矩阵的逆？

(ii)每一个非零的 $n\times n$ 矩阵 A 都有逆矩阵吗？

(iii)如果不是，$n\times n$ 矩阵 A 应满足什么条件才有逆矩阵？

为了回答这些问题需要对矩阵定义某些运算，在实数集中没有类似的运算.

定义 5.15 设 A 是 $m\times n$ 矩阵，A 的初等行(列)运算是对 A 的行(列)进行下列操作：

(i)交换 A 的两行(列).

(ii)用非零数乘 A 的一行(列).

49

(iii) A 的一行(列)乘一个数加到另一行(列).

例 5.16 交换以下矩阵的第1行和第3行：

$$A=\begin{bmatrix}1 & 2\\ 3 & 4\\ 5 & 6\end{bmatrix},$$

得到的矩阵是

$$B=\begin{bmatrix}5 & 6\\ 3 & 4\\ 1 & 2\end{bmatrix}.$$

记 $B=PA$ 和 $A=PB$，其中

$$P=\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{bmatrix}$$

是一个置换阵. 同样的置换运算将 B 回到 A. ■

例 5.17 用$\sqrt{2}$乘以

$$A=\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{bmatrix}$$

的第2行得到

$$B=\begin{bmatrix}1 & 2 & 3\\ 4\sqrt{2} & 5\sqrt{2} & 6\sqrt{2}\end{bmatrix}.$$

注意 $B=DA$ 和 $A=D^{-1}B$，其中

$$D=\begin{bmatrix}1&0\\0&\sqrt{2}\end{bmatrix},\quad D^{-1}=\begin{bmatrix}1&0\\0&\sqrt{2}/2\end{bmatrix}$$

是对角矩阵，与单位阵 I_2 只有一个对角元不同．另一个行的乘法运算将 B 变回到 A．■

例 5.18 矩阵

$$B=\begin{bmatrix}1&2\\-12&-14\\5&6\end{bmatrix}$$

50

是矩阵

$$A=\begin{bmatrix}1&2\\3&4\\5&6\end{bmatrix}$$

的第 2 行加上第 3 行乘以 -3 得到的．若用 3 代替 -3，类似的行运算将 B 变回到 A．该运算相当于左乘一个初等矩阵：$B=EA$ 和 $A=E^{-1}B$，其中

$$E=\begin{bmatrix}1&0&0\\0&1&-3\\0&0&1\end{bmatrix},\quad E^{-1}=\begin{bmatrix}1&0&0\\0&1&3\\0&0&1\end{bmatrix}.$$

这个初等矩阵与单位阵相差一个非对角元．■

一般来说，行(列)初等运算相当于左(右)乘一个可逆矩阵．

注 5.19 对于可逆矩阵 A 线性方程组 $Ax=b$ 有且仅有一个解 $x=A^{-1}b$．一般来说，求解 $Ax=b$ 相当于求 A 的一个最大规模的子矩阵的逆．

注 5.20 有时数称为标量(scalar)以区别于向量，然而数也可看作 1×1 矩阵，因而可视为行向量或列向量．数的加法和乘法与矩阵的加法和乘法一致，但是用矩阵乘法解释矩阵乘以标量是棘手的．一种解释涉及标量矩阵，它是具有相同对角元的方对角阵．

有 n 个分量写成一行或一列的向量可以看作 n 维空间的点，或者从原点出发的矢量(有向线段)．这种表达来自 $n=1, 2, 3$ 的情况，对应于几何与力学中的研究对象．看来矩阵与某些几何变换有关．

用 Irving Kaplansky 关于矩阵的一段引语作为结束(来自 Paul Halmos: *Celebrating* 50 *Years of Mathematics*, New York: Springer-Verlag, 1991)：

> 我们(他和 Halmos)分享关于线性代数的哲理：我们思索无基底，我们书写无基底，但当计算机芯片失效时我们关上办公室的大门，迅速地用矩阵计算．

51

练习

1～8. 设 A 是 1×4 矩阵[1, 2, 0, -3]，B 是 1×4 矩阵[0, -1, -2, 4]，

计算下列各式.

1. $2A+3B$

2. AB^{T}

3. BA^{T}

4. $A^{\mathrm{T}}B$

5. $B^{\mathrm{T}}A$

6. $(A^{\mathrm{T}}B)^2$

7. $(A^{\mathrm{T}}B)^3$

8. $(A^{\mathrm{T}}B)^{1\,000}$

9. 你能计算乘积 AB 或 BA？为什么能，或为什么不能？

10. 找一个方阵 A，使得 $A\neq 0$，但是 $A^2=0$.

11. 找两个方阵 A 和 B，使得矩阵乘积 AB 和 BA 有定义，且 $AB=0$，但 $BA\neq 0$.

12～14. 以矩阵形式书写方程组. 提示：其中有些方程不是线性方程的标准形.

12. $5a+2b+3c-d=1$，$-b+a-3c=2$，$6c+a-3c+1=a$.

13. $x+2b+3c-y=1$，$-b+a-3c=2y$，$x+a-3c+1=a$.

14. $x+2b+3c-y=1$，$-b+a-3c=2y+d$，$x+6a-y+1=a$.

15～17. 解方程组 12～14.

18～25. 对矩阵 $A=[0, 1, -1, 0, 0, 2]$，$B=[-2, 1, 0, 0, 3, 2]$，作练习 1～8.

26～33. 对矩阵 $A=[1, 1, -1, 0, 0, 2, 0]$，$B=[-1, 1, 0, 3, 3, 2, -1]$，作练习 1～8.

34. 对于例 5.6 的矩阵 A，B 作练习 2～7.

35. 对于例 5.7 的矩阵 C 计算 E_1C 和 E_2C，其中 $E_1=\begin{bmatrix}3 & 0\\ 0 & -2\end{bmatrix}$(对角矩阵)，$E_2=\begin{bmatrix}1 & 5\\ 0 & 1\end{bmatrix}$(初等矩阵)，并计算$(E_1)^2$，$(E_1)^9$，$(E_2)^{40}$.

36. 对于例 5.7 的矩阵 C，D 计算 CE_1，CE_2，DE_1，DE_2，E_1E_2，E_2E_1，$(E_1)^2$，$(E_1)^9$，$(E_2)^{40}$，其中 $E_1=\begin{bmatrix}2 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 4\end{bmatrix}$(对角矩阵)，$E_2=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -3 & 0 & 1\end{bmatrix}$(初等矩阵).

52

练习 35～36 的要点是如何实现初等矩阵和对角矩阵(其定义将在下一节提醒)的乘法. 一般来说，作矩阵乘法时矩阵中的零有用.

37. 设 α 是 $m\times m$ 可逆矩阵，β 是 $m\times n$ 矩阵，γ 是 $n\times m$ 矩阵，δ 是 $n\times n$ 矩

阵，计算

$$\begin{bmatrix} I_m & 0 \\ -\gamma\alpha^{-1} & 1_n \end{bmatrix}\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}\begin{bmatrix} I_m & -\alpha^{-1}\beta \\ 0 & 1_n \end{bmatrix}.$$

38. 矩阵 $A=[a_{ij}]$称为对角的，如果 $a_{ij}=0(i\neq j)$. 证明：两个同样大小的对角矩阵的和是对角矩阵；两个对角矩阵的积(当有定义时)是对角矩阵；对于两个 $n\times n$ 对角矩阵 A，B 有 $AB=BA$.

39. 矩阵 $A=[a_{ij}]$称为上三角的，如果 $a_{ij}=0(i>j)$. 证明：两个同样大小的上三角矩阵的和是上三角矩阵；两个上三角矩阵的积(当有定义时)是上三角矩阵；对于某些 $n\times n$ 上三角矩阵 A，B 有 $AB\neq BA$.

40. 矩阵 $A=[a_{ij}]$称为下三角的，如果 $a_{ij}=0(i<j)$. 证明：两个同样大小的下三角矩阵的和是下三角矩阵；两个下三角矩阵的积(当有定义时)是下三角矩阵；证明对于某些 $n\times n$ 下三角矩阵 A，B 有 $AB\neq BA$.

41～44. 用矩阵 A 的行和列的初等变换得到一个对角矩阵. 注：答案中的对角矩阵不是唯一的，但是其中零列的个数是唯一的. 在很多线性代数教科书中证明了这个数等于 $Ax=0$ 的解空间的维数.

41. $A=\begin{bmatrix} 1 & -2 & -1 & 0 \\ 5 & 1 & 3 & 1 \end{bmatrix}$

42. $A=\begin{bmatrix} 1 & 0 & -1 \\ -5 & 1 & 3 \\ 2 & 4 & 5 \\ 8 & -2 & -1 \end{bmatrix}$

43. $A=\begin{bmatrix} 1 & -2 & -1 \\ 5 & 1 & 3 \\ 2 & 4 & 5 \\ -1 & -2 & 0 \\ 8 & -2 & -1 \end{bmatrix}$

44. $A=\begin{bmatrix} 1 & -2 & -1 & 0 \\ 5 & 1 & 3 & 1 \\ 5 & 6 & 0 & 7 \end{bmatrix}$

53

6. 线性方程组

这一节的主要目的是回顾如何解线性方程组，为简便起见我们利用矩阵.

初等运算的主要应用之一是解线性方程组，每个线性方程组都可以用矩阵形式写成 $Ax=b$，其中 A 是给定的矩阵，x 是变量的列(未知)，b 是已知数的列，系数矩阵 A 和 b 可以放在一起形成增广矩阵$[A\mid b]$.

例 6.1 方程组$\begin{cases} x-2z=-3 \\ 2x+5y-4z=5 \end{cases}$ 可以写成

$$\begin{bmatrix}1 & 0 & -2\\2 & 5 & -4\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}-3\\5\end{bmatrix},$$

增广矩阵是$\begin{bmatrix}1 & 0 & -2 & -3\\2 & 5 & -4 & 5\end{bmatrix}$. ■

对增广矩阵的行作乘法运算相当于用非零数乘以方程组的一个方程，行的加法运算相当于把一个方程的若干倍加到另一个方程上，两行交换相当于两个方程交换，两列交换相当于两个变量交换.

对各个方程的这些运算将方程组作等价变换(有相同的变量集合及相同的解集合)，因此，对增广矩阵作初等行运算将其变为简化形式，可以用来解任意的线性方程组. 例如，若系数矩阵是对角的(见上一节练习 38)，则方程组分解为独立的线性方程，每个方程只有一个变量，因而很容易求解.

另一种列的初等运算相当于变量的交换，对于求解系数是数(而不是函数、微分算子等)的线性方程组来说不一定用得到，但是其他情形可能有用. 如果在解线性方程组时用到这种运算，必须保留变量改变的记录，以便在结束时可以用原变量写出最终结果. 为了记下这种改变通常对增广矩阵$[A\mid b]$再附加单位阵I_n，以矩阵

54

$$\begin{bmatrix}A & b\\I_n & \end{bmatrix}$$

开始，其中 n 是变量的个数. 用行和列的加法运算可把系数矩阵变成对角形，然后用乘法运算完成新变量方程组的求解，此时处于附加 I_n 位置的矩阵用于以原变量写出答案.

由于变量的改变会导致附加运算，很多教科书避免这样做，而只作行运算. 每一个系数矩阵都可以通过行运算化成所谓阶梯形式，继而化为一个简化阶梯形，这样容易完成方程组的求解. 阶梯形是带特殊性质的上三角形(见上一节练习 39).

本节我们解释了如何用行初等运算和列交换运算将任意矩阵对角化，以这种方法保持变量集合不变，在具有相同解集合的意义上，从原来到最终的所有线性方程组都是等价的.

著名的高斯-若尔当消去法用向前置换(“向下”的行加法运算)得到一个上三角矩阵 U(有时需要列交换，U 主对角线上所有的零在最下面)，然后用向后置换(“向上”的行加法运算)得到一个对角矩阵(系数矩阵中的零行可能出现某些复杂情况无法得到对角矩阵)，因而这个方法用了列交换，但是没有列的加法或乘法运算. 这种方法的一个解释是我们可以将大小为 $m\times n$ 的系数矩阵 A 写成 $A=LUP$，其中 L 是下三角 $m\times m$ 阵，P 是 $n\times n$ 交换阵，U 是上三角$m\times n$阵.

实际上，可以最少限度地运用列运算及相应的变量交换，方法如下. 用行加

法运算我们总能把任意大小 $m\times n$ 的任何一个系数矩阵化为上三角阵(主对角线下的元全是零)，完成这项工作的步骤是：首先看第 1 列，如果可能的话使它的第 1 个元非零，然后把这个元作为旋转元消去其他所有的元(即将它们变成零)， [55] 继而对去掉第 1 行得到的子矩阵做同样的工作，依此进行下去.

现在分别考虑下面 4 种情形：

情形 1：U 的所有对角元非零，且矩阵是方的(即未知数个数 n 等于方程个数 m).

情形 2：U 的所有对角元非零，且 $n>m$.

情形 3：U 的所有对角元非零，且 $n<m$.

情形 4：U 的一个对角元为零.

对于情形 1，可以用 n 次行乘法运算使所有 n 个对角元等于 1，然后用若干(至多 $n(n-1)/2$)次"向上"的行加法运算使系数矩阵成为 I_n，于是增广矩阵变成 $[I_n \mid b']$，方程组可以求解，且有唯一解 $x=b'$. 事实上，我们只涉及系数矩阵 A 可逆的情况，并且描述了寻求解答 $x=A^{-1}b=b'$ 的方法.

对于情形 2，像上面一样将 U 的前 m 行、m 列构成的子矩阵化为 I_m，增广矩阵变成 $[I_m, c \mid b']$，最终解答是 $y=-cz+b'$，其中 y 是 x 中的前 m 个变量，而其余变量 z 取任意值.

对于情形 3，系数矩阵的后 $m-n$ 行为零，如果增广矩阵的后 $m-n$ 行也为零，那么这些行的意思是 $0=0$，它们是多余的，删掉后简化为情形 1，可以用 n 次行乘法运算和若干次行加法运算求得唯一解. 另一方面，如果增广矩阵最后一列的后 $m-n$ 个元不全为零，则得到形如 0＝非零数的线性方程，因此方程组无解.

最后，对于情形 4，我们先用列交换运算，为了记录它们可在系数矩阵上方写出变量，当交换列时也交换上方相应的标记，用行运算和列交换运算将增广矩阵变为如下形式：

$$\begin{array}{cc|c} y^{\mathrm{T}} & z^{\mathrm{T}} & \\ \end{array}$$
$$\left[\begin{array}{cc|c} U & c & b' \\ 0 & 0 & b'' \end{array}\right],$$

其中 U 是上三角方阵，对角元非零，y 是 x 中标记 U 的那些变量，z 是其余变量，$[0, 0]$代表新的系数矩阵中的一个或几个零行. 如果 $b''\neq 0$(即 b''的一个元非零)，则得到形如 0＝非零数的方程，因此方程组无解. 否则，删掉增广矩阵的零行，简化为情形 2. [56]

这样，线性方程组 $Ax=b$ 的完整解答是下列标准形式之一：

- $0=1$(即方程组无解).
- $x=d$(方程组有唯一解).
- $z=Cy+d$(方程组有无穷多解)，列 z 包括 x 的某些变量，列 y 包括其余

变量.

在最后一种情形，C 是常数 $k\times l$ 矩阵，其中 k 是 z 中变量的个数，l 是 y 中变量的个数，d 是有 k 个元的常数列.

数 k 称为 A 的秩，数 l 称为解集的维数. y 中的变量取任意值，它们决定了 z 中变量的值. 对有唯一解的第 2 种情形，我们说解集的维数是 0.

对于所有 3 种情形，答案中的方程或方程组与原来的 $Ax=b$ 在有完全相同解集(空间)的意义上是等价的，理解这一点是重要的. 可能出现的情况是，$Ax=b$ 可以对不同的变量 y 的子集求解，但是对于所有正确的解答，数 k，l 是不变的.

注 6.2 解线性方程组通常意味着对所有解的完整表述(或说明解不存在)；相反，解优化问题通常意味着求一个最优解和相应的最优值(或说明它们不存在).

注 6.3 一些教科书中方程组 $Ax=b$ 的解以形式 $x=Ct+d$ 给出，其中 t 是含 l 个变量(参数)的列，与 x 中的所有变量不同. 此时并不讨论关于不同变量集合的两个方程组的等价性，而是讨论以某种明确方式给出的两个解空间之间的一一对应. $t\mapsto Ct+d$ 形式的变换称为*仿射变换*，于是解空间(当它非空时)在仿射变换下等同于所有 l 元数组.

57

我们没有借助向量空间和维数的概念描述了求解线性方程组的过程，在这些概念和代数一词出现至少 2000 年之前，人类就在解这种方程组. 顺便指出，这个词的原义是“约化”，它的主要思想是用减少变量个数的方法解方程组.

然而这些概念的确给出了另一种理解. 带有系数 c_1，…，c_l 的向量(比如说是行)v_1，…，v_l 的线性组合是 $c_1v_1+\cdots+c_lv_l$，这些向量被称为线性相关，如果它们中的一个是其他向量的线性组合. 一个矩阵的秩是它线性无关行(或列)数的最大值. 线性方程组有解，当且仅当系数矩阵 A 的秩等于增广矩阵 $[A\mid b]$ 的秩(这是一个很容易的练习). 如果秩相同，则解集的维数等于 A 的列数减去秩.

现在给出几个解方程组的例子.

问题 6.4 在例 6.1 中对 x，y，z 求解线性方程组.

解 从增广矩阵

$$\begin{bmatrix}1 & 0 & -2 & -3\\ 2 & 5 & -4 & 5\end{bmatrix}$$

开始，第 1 行乘以 -2 加到第 2 行(即第 2 行用它减去第 1 行乘 2 代替)，得到上三角阵

$$\begin{bmatrix}1 & 0 & -2 & -3\\ 0 & 5 & 0 & 11\end{bmatrix},$$

有非零对角元 1，5，属于情形 2. 第 2 行乘以 1/5，得到最终矩阵

$$\begin{bmatrix} 1 & 0 & -2 & -3 \\ 0 & 1 & 0 & 11/5 \end{bmatrix}.$$

于是解为：$x=-3+2z$，$y=11/5=2.2$，z 任意．我们没有在系数矩阵上方写变量名，因为没有变量交换，然而当需要记住增广矩阵与方程组的联系时，可以用附加的信息标注矩阵：

58

$$\begin{array}{ccc|c} x & y & z & = \\ \end{array}$$
$$\left[\begin{array}{ccc|c} 1 & 0 & -2 & -3 \\ 0 & 1 & 0 & 11/5 \end{array}\right].$$

■

问题 6.5 解方程组 $x+2y=1$，$3x+6y=2$.

解 增广矩阵为

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 6 & 2 \end{bmatrix}.$$

第 1 行乘以 -3 加到第 2 行，得到上三角阵

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix}.$$

考察第 2 行，结论是方程组无解，写出解答的简短形式是 $0=1$. ■

问题 6.6 解

$$\begin{cases} x+2y=1 \\ 3x+6y=3. \end{cases}$$

解 作与例 6.5 同样的行加法运算，增广矩阵中得到一个零行，删去它，得到最终的增广矩阵为

$$\left[\begin{array}{cc|c} 1 & 2 & 1 \end{array}\right].$$

■

答案 $x=-2y+1$，y 任意.

问题 6.7 对 x 解 $ax=b$，其中 a，b 是给定的数.

答案 若 $a\neq 0$，则 $x=b/a$；若 $a=b=0$，则 x 任意；若 $a=0\neq b$，则无解.

问题 6.8 解方程组

$$\begin{cases} x+2y+3z=4 \\ z=2. \end{cases}$$

解 若对 x，y，z 写出增广矩阵，即

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix},$$

59

属于情形 4. 交换 y 和 z，得到矩阵

$$\begin{array}{ccc} x & z & y \end{array}$$
$$\left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 1 & 0 & 2 \end{array}\right].$$

第 2 行乘以 −3 加到第 1 行，得到最终矩阵：

$$\begin{array}{ccc|c} x & z & y & \\ \end{array}$$
$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & -2 \\ 0 & 1 & 0 & 2 \end{array}\right].$$

现在可以用标准形式写出最终解为

$$\begin{cases} x = -2y - 2 \\ z = 2, \end{cases}$$

其中 y 任意. ■

注意，在这种情况下利用矩阵并没有节约时间和空间，我们用到的加法运算与将第 2 个方程代入第 1 个方程一样. 然而对于有很多非零系数的大型方程组，当利用矩阵去表达信息，又不必写出变量名时，我们会节省时间和空间.

问题 6.9 对 x_1，x_3，x_5 求解

$$\begin{cases} 2x_1 + 3x_2 + x_5 + 5 - 2x_2 = x_1 + x_4 + 1 \\ 2x_3 + x_5 + 2 = -3x_1 - 5x_2 + 3 \\ 3x_3 + 3x_2 - 1 = -x_5 - 6x_4 - 1. \end{cases}$$

解 该线性方程组没有以标准形式给出，将其用标准形式写出，增广矩阵为

$$\begin{array}{ccccc|c} x_1 & x_3 & x_5 & x_2 & x_4 & = \end{array}$$
$$\left[\begin{array}{ccccc|c} 1 & 0 & 1 & 1 & -1 & -4 \\ 3 & 2 & 1 & 5 & 0 & 1 \\ 0 & 3 & 1 & 3 & 6 & 0 \end{array}\right].$$

利用两次向下的加法运算，得到上三角阵：

$$-3\downarrow \begin{array}{c} \\ \leftarrow \\ \rightarrow \end{array} \left[\begin{array}{ccccc|c} 1 & 0 & 1 & 1 & -1 & -4 \\ 3 & 2 & 1 & 5 & 0 & 1 \\ 0 & 3 & 1 & 3 & 6 & 0 \end{array}\right]$$
$$\Downarrow$$
$$-3/2\downarrow \begin{array}{c} \leftarrow \\ \rightarrow \\ \end{array} \left[\begin{array}{ccccc|c} 1 & 0 & 1 & 1 & -1 & -4 \\ 0 & 2 & -2 & 2 & 3 & 13 \\ 0 & 3 & 1 & 3 & 6 & 0 \end{array}\right]$$
$$\Downarrow$$
$$\left[\begin{array}{ccccc|c} 1 & 0 & 1 & 1 & -1 & -4 \\ 0 & 2 & -2 & 2 & 3 & 13 \\ 0 & 0 & 4 & 0 & 1.5 & -39/2 \end{array}\right].$$

60

现在做两次行乘法运算使对角元等于 1：

$$\begin{matrix}\\ 1/2. \\ 1/4.\end{matrix}\begin{bmatrix}1 & 0 & 1 & 1 & -1 & | & -4\\ 0 & 2 & -2 & 2 & 3 & | & 13\\ 0 & 0 & 4 & 0 & 1.5 & | & -39/2\end{bmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix}1 & 0 & 1 & 1 & -1 & | & -4\\ 0 & 1 & -1 & 1 & 3/2 & | & 13/2\\ 0 & 0 & 1 & 0 & 3/8 & | & -39/8\end{bmatrix}.$$

再做两次向上行加法运算(回代)：

$$\begin{matrix}\rightarrow & \rightarrow\\ -1\uparrow & 1\uparrow \rightarrow\\ \leftarrow & \leftarrow\end{matrix}\begin{bmatrix}1 & 0 & 1 & 1 & -1 & | & -4\\ 0 & 1 & -1 & 1 & 3/2 & | & 13/2\\ 0 & 0 & 1 & 0 & 3/8 & | & -39/8\end{bmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix}1 & 0 & 0 & 1 & -11/8 & | & 7/8\\ 0 & 1 & 0 & 1 & 15/8 & | & 13/8\\ 0 & 0 & 1 & 0 & 3/8 & | & -39/8\end{bmatrix}.$$

可将解写成标准形式：

$$\begin{cases}x_1 = -x_2 + 11x_4/8 + 7/8\\ x_3 = -x_2 - 15x_4/8 + 13/8\\ x_5 = -3x_4/8 - 39/8.\end{cases}$$

■

问题 6.10 对 x_1，x_2，x_3 求解与问题 6.9 同样的线性方程组. [61]

解 考察带有列交换的最后的增广矩阵

$$\begin{matrix}x_1 & x_2 & x_3 & x_5 & x_4 & | & =\end{matrix}$$
$$\begin{bmatrix}1 & 1 & 0 & 0 & -11/8 & | & 7/8\\ 0 & 1 & 1 & 0 & 15/8 & | & 13/8\\ 0 & 0 & 0 & 1 & 3/8 & | & -39/8\end{bmatrix}.$$

显然不能以这样的形式写解答：

$$\begin{cases}x_1 = x_4, x_5 \text{ 的一个仿射函数}\\ x_2 = x_4, x_5 \text{ 的一个仿射函数}\\ x_3 = x_4, x_5 \text{ 的一个仿射函数}.\end{cases}$$

那么，怎样确切地表述问题是不清楚的. 然而如果问题是必须用给定的数 x_4，x_5 对未知数 x_1，x_2，x_3 求解方程组，下面是一种解决问题的方法.

若 $x_5+3x_4/8\neq-39/8$，则无解. 否则，删去矩阵的最后一行，交换 x_2 列和 x_3 列，得到最终矩阵

$$\begin{bmatrix} x_1 & x_3 & x_2 & x_5 & x_4 & | & = \\ 1 & 0 & 1 & 0 & -11/8 & | & 7/8 \\ 0 & 1 & 1 & 0 & 15/8 & | & 13/8 \end{bmatrix}.$$

因此，当 $x_5+3x_4/8=-39/8$ 时，标准形式的解是

$$\begin{cases} x_1 = -x_2 + 11x_4/8 + 7/8 \\ x_3 = -x_2 - 15x_4/8 + 13/8, \end{cases}$$

其中 x_2 任意. ■

如果两个方程组有相同的解，称它们是等价的. 当对增广矩阵实施行的初等运算时，我们把这个线性方程组变为一个等价组. 反过来对吗? 见本节最后的练习 40 和 41. 包含不同变量集合的等价性又如何? 我们把这个棘手的问题留给读者作为思考题. 如果不更改变量，也没有消除变量以转换为较少变量的独立方程组，就不需要处理这个问题.

62

我们在第 4 节陈述过下面的事实:

定理 6.11 *给定线性方程组 $Ax=b$，以及来自这个方程组并以标准形表示的另一个方程 $f_0=b_0$，则或者方程 $f_0=b_0$，或者方程 $0=1$ 是该方程组中的方程的一个线性组合.*

证明 注意，如果作一次初等行运算，则新方程组中所有方程都是原来方程的线性组合. 事实上，这些方程中只有一个是新的，在乘法运算情形下它是原来方程的一个线性组合；在加法运算情形下它是两个原来方程的线性组合. 由于线性组合的线性组合仍然是线性组合，所以在任意次初等行运算之后，每个新方程是原来方程的线性组合.

如果原方程组无解，则会得到形如 $0=c(c\neq 0)$的方程，一次乘法运算即可将它变成 $0=1$.

当原方程组恰有一个解 $x=d$ 时用高斯-若尔当方法给出的每个方程 $x_i=d_i$ 都可作为原方程的一个线性组合. 由于 $x=d$ 满足线性方程 $f_0=cx=b_0$，得到 $cd=b_0$，用系数 c_i 把方程$x=d$结合在一起即得 $cx= cd=b_0$.

最后，设原方程组有无穷多解，则高斯-若尔当方法给出形如 $z-Cy=d$ 的方程组，将它代入 $f_0=cx=b_0$ 得到 C，d，c，b 之间的关系. 若记 $f_0=cx= c'y+c''z$，这个关系是 $c'(Cz+d)+c''z=0$，对于所有的 z 成立，于是 $c'C+c''=0$，$c'd=0$. 取方程$z-Cy=d$的线性组合，得到方程 $f_0=b_0$. ■

注 6.12 能够像解线性方程组那样，用依次消去一个变量的方法求解线性不等式组吗? 大约 200 年前傅里叶(Fourier)提出了这个问题，并给予肯定的回答，见附录 A10 中的 Fourier-Motzkin 方法. 然而，没有人能够用其他办法对此方法进行足够的完善，这个方法的问题是约束的个数可能增加得非常快. 下面是

63

消去法能正常运行的一个例子.

问题 6.13 求下面不等式组的一个可行解：

$$\begin{cases} x_1 + x_2 + x_3 + x_4 \leqslant 3 \\ x_1 + 2x_2 + 3x_3 \geqslant 1 \\ \text{所有 } x_i \geqslant 0. \end{cases} \tag{6.14}$$

解 以下列形式写出关于 x_4 的全部两个约束：

$$0 \leqslant x_4 \leqslant 3 - x_1 - x_2 - x_3, \tag{i}$$

从中消去 x_4，得到关于 3 个变量的不等式组

$$\begin{cases} 0 \leqslant 3 - x_1 - x_2 - x_3 \\ x_1 + 2x_2 + 3x_3 \geqslant 1 \\ \text{所有 } x_i \geqslant 0. \end{cases}$$

关于 x_3 的全部 3 个约束为

$$0, (1 - x_1 - 2x_2)/3 \leqslant x_3 \leqslant 3 - x_1 - x_2 \tag{ii}$$

从中消去 x_3，得到关于 2 个变量的不等式组

$$\begin{cases} 0, (1 - x_1 - 2x_2)/3 \leqslant 3 - x_1 - x_2 \\ x_1, x_2 \geqslant 0. \end{cases}$$

最后，为了完成向前的代入以下列形式写出关于 x_2 的全部 3 个约束：

$$0 \leqslant x_2 \leqslant 3 - x_1, 8 - 2x_1 \tag{iii}$$

从中消去 x_2，得到关于 1 个变量的不等式组

$$0 \leqslant 3 - x_1, 8 - 2x_1; x_1 \geqslant 0,$$

它等价于 $0 \leqslant x_1 \leqslant 3$. 取 x_1 任意一个可行的数值，依次求出 x_2，x_3，x_4 的可行的数值. 用这种向后的代入可以得到原不等式组(6.14)的全部可行解.

例如，从 $x_1 = 2$ 开始，对 x_2 的约束(iii)变为 $0 \leqslant x_2 \leqslant 1$. 选取 $x_2 = 0$，约束(ii)为 $0 \leqslant x_3 \leqslant 1$. 取 $x_3 = 0$，约束(i)为 $0 \leqslant x_4 \leqslant 1$. 取 $x_4 = 1$，完成. 于是找到(6.14)的一个可行解. ■

64

注 6.15 在实际生活中，大多数线性方程组用的是有理数数据，因此解也只含有理数，但是解中的有理数可能有很大的分子和分母，计算很费时间. 通常在具有有限精度的计算机上求解方程组，对于实际应用来说近似解常常就可以了. 数据的不确定性给出一个不寻求精确解的好理由.

练习

1～8. 用行和列的加法运算把矩阵 A 化简成对角形，由此确定 A 是否可逆. 注：解答中的对角阵不是唯一的，但是其中对角元的乘积是唯一的. 很多线性代数教科书里证明了这个乘积等于 A 的行列式(determinant).

1. $A = \begin{bmatrix} 1 & 2 \\ 6 & 8 \end{bmatrix}$

2. $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

3. $A=\begin{bmatrix} a & 0 & 0 \\ 0 & b & 1 \\ 0 & 0 & c \end{bmatrix}$

4. $A=\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

5. $A=\begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & -1 & 0 & 1 \\ -1 & 1 & 2 & 1 \end{bmatrix}$

6. $A=\begin{bmatrix} 1 & 0 & -1 \\ 5 & 1 & 3 \\ 2 & 4 & 5 \end{bmatrix}$

7. $A=\begin{bmatrix} 1 & -2 & -1 \\ 5 & 1 & 3 \\ 2 & 4 & 5 \end{bmatrix}$

8. $A=\begin{bmatrix} 1 & -2 & -1 & 0 \\ 5 & 1 & 3 & 1 \\ 0 & 1 & 0 & 1 \\ 5 & 6 & 0 & 7 \end{bmatrix}$

9～16. 对 x，y 求解. 提示：方程可能没有以标准形式给出，b，t，u，z 作为给定数处理.

9. $\begin{cases} x+2y=1 \\ 2x+4y=3 \end{cases}$

10. $\begin{cases} x+2y=3 \\ 2x+4y=6 \end{cases}$

11. $\begin{cases} 2x+3y+5z=2, \\ 3x+5y+8z=b \end{cases}$

12. $\begin{cases} 2x+3y+1=2+x-y, \\ 3x+5y+8=2x \end{cases}$

13. $\begin{cases} x+2y=3+u \\ 2x+4y=t \end{cases}$

65

14. $\begin{cases} x+ty=3 \\ 2x+4y=1 \end{cases}$

15. $\begin{cases} x+t^2y=1 \\ x+y=t \end{cases}$

16. $\begin{cases} x+ty=t^2 \\ 2x+uy=1 \end{cases}$

17. 对 y，z 求解第 11 题的方程组.

18. 对 x，z 求解第 11 题的方程组.

19. 是否存在一个恰好有三个解的方程组?

20～23. 对 5～8 题中的矩阵 A，求 A^{-1}(如果存在).

24～27. 如果可能的话，把 1～4 题中的矩阵 A 写成 $A=LU$，其中 U 是一个上三角矩阵，L 是一个下三角矩阵.

28～31. 如果可能的话，把 5～8 题中的矩阵 A 写成 $A=LU$，其中 U 是一个上三角矩阵，L 是一个下三角矩阵. 然后计算矩阵 UL.

32～35. 对 x，y，z 求解方程组.

32. $\begin{cases} x+y+z=a+b^2, \\ x+2y+3z=c^3, \\ x+3y+4z=d. \end{cases}$

33. $\begin{cases} x+5y+z=y, \\ x+2y+3=z, \\ x+3y-4z=d. \end{cases}$

34. $\begin{cases} x+y+z=u_1, \\ x+2y+3z=u_2, \\ x+3y+4z=u_3. \end{cases}$

35. $\begin{cases} u+y+z=x, \\ x+2u+3z=y, \\ x+3y+4z=v. \end{cases}$

36～39. 找出下面给出的系统的一个可行解(或证明不存在可行解).

36. x，$y\geqslant 0$，$x+y\leqslant 2$，$z=3x-4y+5$.

37. x，y，$z\geqslant 0$，$x+y+z\leqslant 2$，$x+2y+3z\geqslant 3$.

38. x，y，$z\geqslant 0$，$x+y-z\leqslant 2$，$x+2y+3z\geqslant 3$.

39. $x-2y+3z\geqslant 5$，$3x+4y-z\leqslant 8$，$x-2y\geqslant -3$.

40. 给定两个线性方程组 $Ax=b$ 和 $A'x=b'$，其中 x 有 n 个不同的分量，它们有相同的非空解集，并且都有 m 个方程. 证明可以通过行加法运算和行乘法

运算，从增广矩阵$[A \mid b]$得到$[A' \mid b']$. 特别地，存在一个 $m\times m$ 的可逆矩阵 C，使得$CA=A'$，$Cb=b'$.

41. 给定两个线性方程组 $Ax=b$ 和 $A'x=b'$，其中 x 有 n 个不同的分量，它们有相同的非空解集，并且方程的个数 $m>m'$. 证明可以通过行加法运算，从增广矩阵$[A \mid b]$得到 $m\times n$ 的矩阵$\begin{bmatrix} A' & b' \\ 0 & 0 \end{bmatrix}$，其中最后的 $m-m'$行是零行. 特别地，存在一个 $m\times m'$的矩阵 C，使得 $CA=A'$，$Cb=b'$. 证明存在一个 $m'\times m$ 的矩阵 C'，使得 $C'A'=A$，$C'b'=b$.

66

第3章 表与旋转

7. 线性规划的标准型和典范型

为简便起见我们将用矩阵记号表述线性规划. n 个变量 x_1, …, x_n 的线性型 $c_1x_1+c_2x_2+\cdots+c_nx_n$ 可以记作矩阵的乘积 cx，其中 $c=[c_1, \cdots, c_n]$是系数行，$x=[x_1, \cdots, x_n]^T$ 是变量列. 如果从问题本身可以知道变量集合由 x_1, x_2, …,x_n 给出，则线性型由行 c 唯一确定，即若 $z=cx$，$\hat{z}=\hat{c}x$，则当且仅当 $c=\hat{c}$时 $z=\hat{z}$对所有 x 成立.

在第 1 章第 1 节曾经定义过线性约束，可以用标准型记作

$$ax=b，或 ax\leqslant b，或 ax\geqslant b，$$

其中 ax 是 x 的元素的线性型.

对于同样类型的一组不等式，可以用矩阵形式将它们记作 $Ax\leqslant b$ 或 $Ax\geqslant b$. 验证上述说法成立是值得做的练习. 注意，对于同样长度的两个列向量 b, b', $b\leqslant b'$是指 b 的每一个分量都小于等于 b'的对应分量. 对于行向量有类似的约定. 若 b 是一个列(或行)向量，$b\geqslant 0$ 是指 b 的所有分量非负.

例 线性约束集合

$$\begin{cases}0.3x_a+0.35x_b+0.5x_c+0.4x_d\geqslant 0.4\\ 0.6x_a+0.35x_b+0.5x_c+0.45x_d\geqslant 0.5\\ 0.1x_a+0.3x_b+0.15x_d\geqslant 0.1\end{cases}$$

可以写成如下矩阵形式：

$$\begin{bmatrix}0.3 & 0.35 & 0.5 & 0.4\\ 0.6 & 0.35 & 0.5 & 0.45\\ 0.1 & 0.3 & 0 & 0.15\end{bmatrix}\begin{bmatrix}x_a\\ x_b\\ x_c\\ x_d\end{bmatrix}\geqslant\begin{bmatrix}0.4\\ 0.5\\ 0.1\end{bmatrix}.$$

■

线性规划的框架如下给定：

在有限多个线性约束下求一个仿射函数(一个线性型加一个常数)的最大值(或最小值).

本节的主要目的是建立书写线性规划的统一形式，以便任何的线性规划都能以统一形式写出. 当我们要给出求解线性规划的方法，或者叙述线性规划的定理时，就容易看出这种统一性的重要意义. 没有它，就要考虑书写线性规划的所有可能的方式，使得叙述解法或证明非常麻烦. 求解线性规划经常要用到计算机，

所以将线性规划转换成一个给定的软件所能接受的形式是很重要的.

文献中统一形式的不同模型通常称为标准型(standard form)、典范型(canonical form)和正规型(normal form). 在不同出版物中它们不一定指同样的东西. 本节我们首先定义线性规划的标准型和典范型，然后将会看到任何一个线性规划怎样写成典范型和标准型.

定义 7.1 如果一个最小化问题的所有变量都是非负的，所有其他约束都是线性方程，则称这个线性规划为标准型.

换言之，利用矩阵符号，标准型为

$$\text{minimize} \quad cx+d, \text{ subject to } Ax=b, \ x\geqslant 0,$$

其中 c 是给定的行，d 是给定的数，A 是给定的矩阵，x 是变量列，b 是给定的列.

可以将上述式子简记作

$$cx+d \to \min, \ Ax=b, \ x\geqslant 0.$$

注意，这里 $Ax=b$ 是线性方程组的标准型.

例 $2x+3y-z\to\min$，x，y，$z\geqslant 0$，$x+y+z=1$ 是线性规划标准型. ■

定义 7.2 线性规划的典范型是：minimize $cx+d$，subject to $Ax\leqslant b$，$x\geqslant 0$，其中 A，b，c，d，x 同前.

换言之，这是最小化问题，要求所有变量非负，其他约束为$\leqslant$形式.

例 $2x+3y-z\to\min$，x，y，$z\geqslant 0$，$z\leqslant 2$ 是线性规划典范型. ■

在不同的线性规划书中，线性规划标准型、典范型(以及正规型)可能不一样. 但符号约束 $x\geqslant 0$ 对于这些形式是相同的，但是有些书里用 max 代替 min，或者(和)用 $Ax\geqslant b$ 代替 $Ax\leqslant b$ 或 $Ax=b$. 不同的软件包也有不同的线性规划输入形式，有些书和软件在目标函数中不允许有常数 d.

事实上，重要的是知道如何从一种形式转换成另一种形式. 我们将给出这种转换的一些小技巧.

利用初等代数能够将任何一个线性规划写成标准型及典范型. 我们要用到在下面引理中概述的基本事实.

引理 7.3 1. 设 a，b 是实数，则

(a) $a\leqslant b$ 当且仅当 $-a\geqslant -b$.

(b) $a=b$ 当且仅当 $a\leqslant b$ 且 $-a\leqslant -b$.

2. 函数 f 的最小化等价于同样约束下函数 $-f$ 的最大化，即两个优化问题有相同的可行解和相同的最优解，最优值只差一个符号(f 的最小值等于 $-f$ 最大值的负数).

3. 函数 f 的最小化等价于同样约束下函数 $f+d$ 的最小化，即两个优化问题有相同的可行解和相同的最优解，最优值只差 d($f+d$ 的最小值等于 f 的最小值加 d).

证明

(1. a) $a \leqslant b$ 当且仅当 $a-b \leqslant 0$ 当且仅当 $-(a-b) \geqslant 0$ 当且仅当 $b-a \geqslant 0$ 当且仅当 $-a \geqslant -b$.

(1. b) 若 $a=b$，则一定有 $a \leqslant b$，并且 $-a=-b$，而由此有 $-a \leqslant -b$. 反之，由 (1. a) $a \geqslant b$ 和 $-a \leqslant -b$ 等价，于是 $a \leqslant b$ 且 $a \geqslant b$，由此得 $a=b$.

2. 设 c 是 f 达到最小值的点，则对任意 x，$f(c) \leqslant f(x)$. 利用 (1. a)，对任意 x 有 $-f(c) \geqslant -f(x)$，这表示 $-f$ 在 c 点达到最大值.

3. 留给读者作为练习. ■ 69

根据这个引理，我们可以利用下面的“技巧”将任何线性规划问题重写成它的典范型或标准型(这里 f 表示一个线性型，c 是一个常数).

技巧 7.4 根据引理 7.3 的 (1. a)，不等式 $f \leqslant c$ 与 $-f \geqslant -c$ 等价.

由此，可以将线性规划中某些或全部形式为 $f \geqslant c$ 的线性不等式写成 $f \leqslant c$ 的形式. 同时，如果愿意的话，也可以将它们改写成 $f \geqslant c$ 的形式.

技巧 7.5 根据引理 7.3 的 (1. b)，等式 $f=c$ 与不等式组 $f \leqslant c$，$-f \leqslant -c$ 等价.

于是可以用两个不等式代替任意的线性方程. 不失一般性，假定线性规划问题中的所有线性约束都具有 $f \leqslant c$ 或 $-f \leqslant -c$ 的形式.

技巧 7.6 根据引理 7.3 的 (2)，可以用目标函数乘以 -1 的办法将最小化问题转换为最大化问题，反过来也一样.

切记最优值应改变符号!

利用技巧 7.4～7.6，可以将所有变量 $\geqslant 0$ 的任何线性规划以典范型写出，而不改变其变量集合、可行域和最优值. 也就是说，如果是最大化问题，用技巧 7.6 转换为最小化问题，用技巧 7.5 将等式约束(如果有的话)转换为不等式约束，最后用技巧 7.4 将符号限制(指所有变量 $\geqslant 0$)以外的 $\geqslant$ 约束(如果有的话)转换为 $\leqslant$ 约束. 特别地，标准型 $cx+d \to \min$，$Ax=b$，$x \geqslant 0$ 可以转换为等价的典范型：$cx+d \to \min$，$Ax \leqslant b$，$-Ax \leqslant -b$，$x \geqslant 0$.

技巧 7.7 一个约束 $f \leqslant c$ 可以用附加(松弛)变量 s 写成两个约束 $f+s=c$，$s \geqslant 0$.

这样，我们可以将任意一个不等式约束转换为一对约束，一个是等式，另一个是符号限制. 松弛变量 s 有别于问题中的其他变量. 这个技巧也能够反过来用，此时节省(消去)一个变量 s，s 应从所有约束及目标函数中消去，可以得到少一个变量的问题.

技巧 7.7 与 7.4 结合起来用于 $\geqslant$ 约束：$f \geqslant c$ 可以写成 $f-s=c$，$s \geqslant 0$. 这时新变量 s 称为剩余变量(surplus or excess variable)，我们也称它为松弛变量. 70

利用技巧 7.7(如果需要的话，还有技巧 7.6)可以将所有变量 $\geqslant 0$ 的任何 LP 以标准型写出，例如典范型 $cx+d \to \min$，$Ax \leqslant b$，$x \geqslant 0$ 可以写成标准型

$cx+d\to\min$，$Ax+u=b$，$x\geqslant 0$，$u\geqslant 0$，其中 u 是松弛变量的列($Ax\leqslant b$ 中的每个不等式有一个松弛变量)．这个标准型是特殊的，因为从其线性方程组解得：$u=-Ax+b$ 是解的标准形式．

相反地，给定任何一个以标准型 $cx+d\to\min$，$Ax=b$，$x\geqslant 0$ 表示的 LP，可以用解线性方程组 $Ax=b$ 的办法转化为典范型．(如果方程组不相容，则该规划不可行，从而不需要把它写成任何形式．)按照这种办法我们减少约束和变量的个数，而不是用技巧 7.5 去增加约束的个数．

按照技巧 7.7，增加松弛变量或消去某些变量的办法可以改变原规划中的变量集合，这样我们不能说可行域没有变化，然而即使在反复利用这个技巧之后，这种变化也是简单、明确的：原规划 P 的可行域 S 和新规划 P$'$的可行域 S'在仿射变换

$$x=Cx'+B,x'=C'x+B'$$

下是一一对应的，其中 x 是 P 的 n 个变量的列，x'是 P$'$的 n'个变量的列，C，B，C'，B'分别是 $n\times n'$，$n\times 1$，$n'\times n$，$n'\times 1$ 常数矩阵，并且最优域相互转换．于是我们称这两个规划是仿射等价的．

例如，上面的典范型 $cx+d\to\min$，$Ax\leqslant b$，$x\geqslant 0$ 和标准型 $cx+d\to\min$，$Ax+u=b$，$x\geqslant 0$，$u\geqslant 0$ 由仿射变换

$$x=[I_n\quad 0]\begin{bmatrix}x\\u\end{bmatrix},\begin{bmatrix}x\\u\end{bmatrix}=\begin{bmatrix}I_n\\-A\end{bmatrix}x+\begin{bmatrix}0\\b\end{bmatrix}$$

相联系，其中 n 是列 x 的元的个数，0 代表相应大小的零矩阵．

技巧 7.8 可以将线性规划中的任意变量 x 写成两个附加(人工)变量 x'，x''之差 $x'-x''$，且 $x'\geqslant 0$，$x''\geqslant 0$．

71

这个技巧的目的是去掉一个可能为负的变量．若已知 $x\leqslant a$，则可用更简单的代换 $x=a-x''$，$x''\geqslant 0$．若已知 $x\geqslant a$，则代换 $x=x'+a$，$x'\geqslant 0$ 成立．若不知道 x 的任何界限，但 x(带非零系数)出现在一个线性方程中，我们可以从这个方程解出 x，并从线性规划与该方程一起将 x 完全消掉．否则，虽然仍可用 Fourier-Motzkin 消去法(见第 6 节末尾)从约束中消掉 x，但是若 x 出现在很多约束中，那么这个方法不如技巧 7.8 实用．若 x 不是在约束而是在目标函数中出现，则这个线性问题不可行或无界．

技巧 7.8 甚至比技巧 7.7 更差，因为它不仅改变了变量集合，而且失去了仿射等价，虽然变换 $x=x'-x''$是仿射的，但是反向的仿射变换不存在，除非 x 有界(这时可以固定 x'或 x'')．有一个反向变换是

$$[x',x'']=\begin{cases}[x,0] & \text{当 } x\geqslant 0\\ [0,-x] & \text{当 } x\leqslant 0,\end{cases}$$

但它不是仿射变换．

利用上述 5 种技巧若干次可以将线性规划 P 变成另一个线性规划 P$'$，它们有

如下的关系：

- 由 P 的每一个可行解容易给出 P′的一个可行解，反之亦成立.
- 由 P 的每一个最优解容易给出 P′的一个最优解，反之亦成立.

将技巧 7.6、7.7 和 7.8 结合起来可得

任何线性规划可以写成标准型.

将技巧 7.4、7.5、7.6 和 7.8 结合起来可得

任何线性规划可以写成典范型.

在某些书和软件中一个线性规划的目标函数必须是线性的，这可以用引理 7.3(3)来处理. 不要忘记这样做会改变最优值！有时计算机软件为我们完成某些技巧. 当问题的规模无关紧要时，即使是只增加变量和约束个数的技巧，也广泛应用在线性规划软件中. [72]

目标函数中允许有常数项，因为即使开始时没有这一项，它也可能在用单纯形法和其他方法求解问题的过程中出现. 下面是几个特殊的例子.

问题 7.9 将 $2x+3y-z\to\min$，x，y，$z\geqslant0$，$x+y+z=1$ 转换为典范型.

解 用前面那些技巧得到典范型 $2x+3y-z\to\min$，x，y，$z\geqslant0$，$x+y+z\leqslant1$，$-x-y-z\leqslant-1$. ■

解决问题 7.9 的另一种方法是，比如对 z 解方程得 $z=1-x-y$. 代入线性规划得到一个更简单的典范型 $x+2y-1\to\min$，x，$y\geqslant0$，$x+y\leqslant1$，容易解出在 $x=0$，$y=0$，$z=1$ 处 $\min=-1$.

问题 7.10 将 $f=2x+3y-z\to\max$，x，$y\geqslant0$，$x+y+z\geqslant1$ 转换为典范型.

解 用 -1 乘目标函数 f 和约束 $x+y+z\geqslant1$，得 $-f=-2x-3y+z\to\min$，x，$y\geqslant0$，$-x-y-z\leqslant-1$. 设 $z=z'-z''$，z'，$z''\geqslant0$，得典范型 $-f=-2x-3y+z'-z''\to\min$，x，$y\geqslant0$，$-x-y-z'+z''\leqslant-1$. ■

注意，$z=z'-z''$ 不是原问题的一部分也不是典范型，而是将它们联系起来的形式. 一般说来，在改变目标函数之前应对它命名. 在需要解一个线性规划时，最终答案应按原问题的要求给出.

用一种或另一种形式重写一个数学问题可能有助于它的求解，或者为利用更完善的数学工具或计算机处理它做准备. 例如，为了对 x 解方程 $4+x-5x=3-x$，先用加减法把它变成标准型 $-3x=-1$，然后用除法求解.

问题 7.11 将食谱问题(例 2.1)写成典范型和标准型.

解 为得到典范型，需要做的是用 -1 乘 3 个形式为 $\geqslant$ 的约束：

$$\begin{cases}10a+15b+5c+60d+8e\to\min\\-0.3a-1.2b-0.7c-3.5d-5.5e\leqslant-50\\-73a-96b-20\,253c-890d-279e\leqslant-4\,000\\-9.6a-7b-19c-57d-22e\leqslant-1\,000\\a,b,c,d,e\geqslant0.\end{cases}$$

[73]

为得到标准型，将 3 个松弛变量加到对应于蛋白质、维生素 A 和钙的约束上：

$$\begin{cases}10a+15b+5c+60d+8e\to\min\\-0.3a-1.2b-0.7c-3.5d-5.5e+u_1=-50\\-73a-96b-20\,253c-890d-279e+u_2=-4\,000\\-9.6a-7b-19c-57d-22e+u_3=-1\,000\\a,b,c,d,e,u_1,u_2,u_3\geqslant 0.\end{cases}$$

■

问题 7.12 将混合问题(例 2.2)写成典范型和标准型.

解 这个问题已经有一个标准型：

$$\begin{cases}1.2a+1.4b+1.7c+1.9d\to\min\\0.9a+0.8b+0.7c+0.6d=0.75\\0.1a+0.2b+0.3c+0.4d=0.25\\a,b,c,d\geqslant 0.\end{cases}$$

我们不必写入多余约束 $a+b+c+d=1$. 利用矩阵可将问题写成

$$[1.2,1.4,1.7,1.9]\begin{bmatrix}a\\b\\c\\d\end{bmatrix}\to\min,$$

$$\begin{bmatrix}0.9&0.8&0.7&0.6\\0.1&0.2&0.3&0.4\end{bmatrix}\begin{bmatrix}a\\b\\c\\d\end{bmatrix}=\begin{bmatrix}0.75\\0.25\end{bmatrix}$$

$$[a,b,c,d]\geqslant 0.$$

为得到典范型将每个方程用两个不等式代替：

$$\begin{cases}1.2a+1.4b+1.7c+1.9d\to\min\\0.9a+0.8b+0.7c+0.6d\leqslant 0.75\\-0.9a-0.8b-0.7c-0.6d\leqslant -0.75\\0.1a+0.2b+0.3c+0.4d\leqslant 0.25\\-0.1a-0.2b-0.3c-0.4d\leqslant -0.25\\a,b,c,d\geqslant 0.\end{cases}$$

■

[74] 得到典范型的另一种办法是对 a，b 解两个线性方程，并从线性规划中消去 a，b，得到只含两个变量 c，d 的典范型. 这个线性规划容易用图解法求解.

注 解线性方程组的高斯-若尔当消去法(见前面第 6 节)也可以看作是利用解集之间一一对应的仿射变换，从一个方程组到含有较少变量的方程组的简化.

而在对于不等式组的 Fourier-Motzkin 消去法中(见第 6 节)涉及一些不等式组之间更复杂的关系. 第 7 章将用到一种非仿射的变量代换，

一些近代方法(如 Karmirkar 方法)使用非仿射变换的变换.

练习

1～11. 将下面的优化问题重写成线性规划标准型和线性规划典范型，答案可以用矩阵形式给出. 不需要求解这些线性规划. 提示：为使结果的规模较小，可以解线性方程消去变量而不是添加新变量.

1. $\begin{cases} \text{maximize} & 2x+3y \\ \text{subject to} & x\geqslant 1,\ y\geqslant -1,\ x+y\leqslant 5. \end{cases}$

2. $x\to\max$，$y=x+1$，$x+y\leqslant 9$，$y\geqslant 1$

3. $x_1-5x_2+x_3+3x_4\to\min$，

   $2x_2+x_3+3x_4=3$，$x_1-x_2+3x_4\geqslant 3$，所有 $x_i\geqslant 0$

4. $x_1-5x_2+x_3+3x_4\to\max$，

   $2x_2+x_3+3x_4=3$，$x_1-x_2+3x_4\geqslant 3$，x_1，x_2，$x_3\geqslant 0$

5～8. 例 1.9～1.12 中的线性规划.

9. $\begin{bmatrix} 1 & x & 2 \\ 0 & 5 & -1 \end{bmatrix}\begin{bmatrix} y \\ -1 \\ x+z \end{bmatrix}=\begin{bmatrix} x \\ 0 \end{bmatrix}$，$\begin{bmatrix} x \\ y \\ z \end{bmatrix}\geqslant 0$，$x+y+z\to\min$

10. $\begin{bmatrix} 1 & -2 & 2 \\ 0 & 5 & -1 \end{bmatrix}\begin{bmatrix} y \\ -1 \end{bmatrix}\geqslant\begin{bmatrix} x \\ 0 \end{bmatrix}$，$\begin{bmatrix} x \\ y \\ z \end{bmatrix}\geqslant 0$，$x+y+z\to\min$

11. $3x_1-x_2+x_3+3x_4+x_5-5x_6+x_7+3x_8+x_9\to\min$，

    $x_1-5x_2+2x_3+2x_4-x_5-x_6-2x_7+3x_8+x_9\geqslant 3$，

    $-x_1+x_2+x_3+x_4+x_5-x_6-2x_7+3x_8+x_9\leqslant -1$，

    $2x_1-2x_2-2x_3+2x_4+3x_5-x_6-2x_7+x_8+x_9=2$，

    $x_1+3x_5-x_6-2x_7-x_9=0$，所有 $x_i\geqslant 0$

75

8. 旋转表

考察下面的线性方程组：

$$\begin{cases} 2x_1+3x_2=4 \\ 5x_1+6x_2=7, \end{cases} \tag{8.1}$$

它可用矩阵记号写成标准型 $Ax=b$：

$$\begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} 4 \\ 7 \end{bmatrix}.$$

这个方程组利用行表(row tableaux)可以重写成

$$
\begin{array}{cc} x_1 & x_2 \end{array}
$$
$$
\begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix} \begin{array}{l} = 4 \\ = 7 \end{array} \qquad (8.2)
$$

或

$$
\begin{array}{ccc} x_1 & x_2 & -1 \end{array}
$$
$$
\begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{bmatrix} \begin{array}{l} = 0 \\ = 0 \end{array}
$$

这两个表中的行对应于线性方程，第 1 个表中的矩阵是系数矩阵 A，第 2 个表中的矩阵是增广矩阵$[A \mid b]$. 我们在两个表的上边缘写变量名.

这样，表(tableau)是一个在边缘上用附加信息做了标志的矩阵，例如边缘上的信息是变量名和常数.

下面用列表写同样的方程组，它用列来表示线性方程，并把变量名写在左边缘：

$$
\begin{array}{c} x_1 \\ x_2 \end{array} \begin{bmatrix} 2 & 5 \\ 3 & 6 \end{bmatrix} \quad \text{或} \quad \begin{array}{c} x_1 \\ x_2 \\ 1 \end{array} \begin{bmatrix} 2 & 5 \\ 3 & 6 \\ -4 & -7 \end{bmatrix}
$$
$$
\begin{array}{cc} =4 & =7 \end{array} \qquad\qquad\qquad \begin{array}{cc} =0 & =0 \end{array}
$$

正如我们所看到的，这些表给出了线性方程组的另一种表示形式，它们也可用于写出和处理线性规划的数据，应用表的主要优点是变量只需在表的边缘写出一次. 表是处理线性规划的一种简便方式，可以节省书写、纸张和时间，习惯了以后也容易阅读.

76

在这个方面表非常类似于用来解线性方程组的矩阵，系数矩阵的列对应于变量，我们不需要在上边缘写出变量名，除非交换变量或者进行与变量改变相对应的列运算. 线性规划中把字母放在矩阵的边缘上以避免混淆.

我们将用表来解释单纯形方法运行的方式. 表可以用于用手求解变量或约束个数较少的线性规划. 当数据矩阵稀疏(即有很多零元)时其他处理数据的方法比表更好.

例 8.3 线性规划的标准型(见定义 7.1)可以用如下形式写成一个行表：

$$
\begin{array}{c} x^{\mathrm{T}} \end{array}
$$
$$
\begin{bmatrix} A \\ c \end{bmatrix} \begin{array}{l} = b \\ \rightarrow \min, \quad x \geqslant 0 \end{array}
$$

其中 A 是系数矩阵，b 是列矩阵，c 是行矩阵. 注意，上边缘的变量写成行矩阵(x^{T} 记变量列矩阵的转置). 目标函数的系数 c 写在表的最后一行，非负约束 $x \geqslant 0$ 写在表外. ■

例 8.4 线性规划典范型(见定义 7.2)可以用如下形式写成一个表：

$$\begin{array}{r} x \\ 1 \end{array}\begin{bmatrix} -A^{\mathrm{T}} & c^{\mathrm{T}} \\ b^{\mathrm{T}} & d \end{bmatrix}$$
$$\downarrow$$
$$\geqslant 0 \quad \min, x \geqslant 0, y \geqslant 0$$

或

$$\begin{array}{r} x \\ 1 \end{array}\begin{bmatrix} -A^{\mathrm{T}} & c^{\mathrm{T}} \\ b^{\mathrm{T}} & d \end{bmatrix}$$
$$\downarrow$$
$$= y \quad \min, x \geqslant 0, y \geqslant 0$$

77

这里我们用了列表，还引入了新的(松弛)变量 y. ■

高斯-若尔当消去法中的一个计算步骤可以解释如下. 为了解线性方程组，首先对于其中一个变量(旋转变量)解一个方程(旋转方程)，然后用代入法从其他方程中消去这个变量.

这样得到少了一个变量和一个方程的较小的方程组，重复这些步骤(向前代入)直到变量和方程都不剩下，更详细的说明见第 6 节.

现在解释这一步——*旋转步*(pivot step)如何用表来进行. 我们从一个小的例子——线性方程组(8.1)和它的行表表示(8.2)开始.

高斯-若尔当消去法中的一步如下进行：

1. 对于一个未知数如 x_1 解方程组中的一个方程 $2x_1+3x_2=4$：

$$x_1 = 4/2 - (3/2)x_2.$$

2. 将 x_1 的表达式代入第 2 个方程 $5x_1+6x_2=7$，得到

$$5(4/2-(3/2)x_2)+6x_2 = 7 \text{ 或 } 5\cdot 4/2+(6-5\cdot 3/2)x_2 = 7.$$

新的表为

$$\begin{array}{cc} 4 & x_2 \end{array}$$
$$\begin{bmatrix} 1/2 & -3/2 \\ 5/2 & 6-5\cdot 3/2 \end{bmatrix}\begin{array}{l} = x_1 \\ = 7 \end{array} \tag{8.5}$$

将(8.1)和(8.2)中的数 2，3，4，5，6，7 换成任意数 α，β，u，γ，δ，v，并将变量 x_1，x_2 重新命名为 x，y. 也就是说，我们考察对 x，y 的两个线性方程的任意方程组

$$\begin{cases} \alpha x+\beta y = u \\ \gamma x+\delta y = v \end{cases} \tag{8.6}$$

和相应的表

$$\begin{array}{cc} x & y \end{array}$$
$$\begin{bmatrix} \alpha^* & \beta \\ \gamma & \delta \end{bmatrix}\begin{array}{l} = u \\ = v \end{array} \tag{8.7}$$

假定带星号标记的元 α 非零，则称之为*旋转元*(pivot entry)，像上面一样对

[78] x 解第 1 个方程并代入第 2 个方程，得到表

$$\begin{array}{cc} u & y \\ \begin{bmatrix} 1/\alpha & -\beta/\alpha \\ \gamma/\alpha & \delta-\beta\gamma/\alpha \end{bmatrix} & \begin{array}{l} =x \\ =v \end{array} \end{array} \tag{8.8}$$

从表(8.7)变到表(8.8)称为旋转步(pivot step)，下面是 2×2 表的旋转步：

$$\begin{array}{cc} x & y \\ \begin{bmatrix} \alpha^* & \beta \\ \gamma & \delta \end{bmatrix} & \begin{array}{l} =u \\ =v \end{array} \end{array} \mapsto \begin{array}{cc} u & y \\ \begin{bmatrix} 1/\alpha & -\beta/\alpha \\ \gamma/\alpha & \delta-\beta\gamma/\alpha \end{bmatrix} & \begin{array}{l} =x \\ =v \end{array} \end{array} \tag{8.9}$$

任意大小的表的旋转类似地进行，下面是更小的表的例子：

$$\overset{x}{[\alpha^*]} = u \rightarrow \overset{u}{[1/\alpha]} = x,$$

$$\begin{array}{cc} x & y \\ [\alpha^* & \beta] \end{array} = u \mapsto \begin{array}{cc} u & y \\ [1/\alpha & -\beta/\alpha] \end{array} = x.$$

下面是更大的表的例子：

$$\begin{array}{cc} x & y \\ \begin{bmatrix} \gamma_1 & \delta_1 \\ \alpha^* & \beta \\ \gamma_2 & \delta_2 \end{bmatrix} & \begin{array}{l} =v_1 \\ =u \\ =v_2 \end{array} \end{array} \mapsto \begin{array}{cc} u & y \\ \begin{bmatrix} \gamma_1/\alpha & \delta_1-\beta\gamma_1/\alpha \\ 1/\alpha & -\beta/\alpha \\ \gamma_2/\alpha & \delta_2-\beta\gamma_2/\alpha \end{bmatrix} & \begin{array}{l} =v_1 \\ =x \\ =v_2 \end{array} \end{array}$$

这些例子表明对于任意大小的表如下进行旋转：

旋转规则

- 交换旋转行和列的字母($x \leftrightarrow u$，这是表的上边缘和右边缘仅有的改变).
- 将旋转元 $\alpha \neq 0$ 用它的倒数 $\alpha' = 1/\alpha$ 代替($\alpha \mapsto 1/\alpha$).
- 旋转行中不在旋转列的每个元 β 除以 $-\alpha$(即每个这样的 β 代之以 $\beta' = -\alpha'\beta$).
- 旋转行和旋转列之外的每个元 δ 代之以 $\delta - \beta\gamma/\alpha = \delta + \beta'\gamma$，其中 β 为旋转行中与 δ 位于同一列(δ 之上或之下)的元，γ 为旋转列中与 δ 位于同一行(δ 之左或之右)的元.
- 旋转列中不在旋转行的每个元 γ 除以旋转元 α(即每个这样的 γ 代之以 $\gamma' = -\alpha'\gamma$).

[79]

于是 $m \times n$ 表的一步旋转经过 1 次字母交换、1 次除法、$mn-1$ 次乘法、$mn-m-n+1$ 次加法和 $n-1$ 次改变符号来完成.

作为练习，我们用 * 号指定表(8.5)中第 2 行第 2 列的元为旋转元，作一步旋转：

$$7 \leftrightarrow x_2$$

$$\alpha = -3/2 \mapsto \alpha' = 1/\alpha = -2/3$$

$$\beta = 5/2 \mapsto \beta' = -\beta/\alpha = -\alpha'\beta = 5/3$$

$$\delta = 1/2 \mapsto = \delta - \beta\gamma/\alpha$$

$$= \delta + \beta'\gamma = 1/2 + (5/3)(-3/2) = 1/2 - 5/2 = -2$$

$$\gamma \mapsto \gamma/\alpha = \gamma\alpha' = -1/17$$

$$\begin{matrix} 4 & x_2 \\ \end{matrix} \qquad\qquad \begin{matrix} 4 & 7 \end{matrix}$$

$$\begin{bmatrix} 1/2 & -3/2 \\ 5/2 & -3/2^* \end{bmatrix} \begin{matrix} = x_1 \\ = 7 \end{matrix} \quad \mapsto \quad \begin{bmatrix} -2 & 1 \\ 5/3 & -2/3 \end{bmatrix} \begin{matrix} = x_1 \\ = x_2 \end{matrix}$$

于是解出方程组(8.1)：

$$x_1 = 4 \cdot (-2) + 7.1 = -1,$$

$$x_2 = 4 \cdot (5/3) + 7 \cdot (-2/3) = 2.$$

事实上，我们求解了一个用任意数 u，v 代替常数项 4，7 的更一般的方程组，其解为

$$x_1 = -2u + v,$$

$$x_2 = 5u/3 - 2v/3.$$

换言之，我们计算了

$$\begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}^{-1} = \begin{bmatrix} -2 & 1 \\ 5/3 & -2/3 \end{bmatrix}.$$

现在对任意大小的表检验旋转规则. 考察由一个行表给出的线性方程组 $AX=b$：

80

$$\begin{matrix} X^{\mathrm{T}} \\ [A] = b \end{matrix}$$

其中 $X^{\mathrm{T}}=[x_1, \cdots, x_n]$ 是变量行，b 是常数列(也可看作变量)，$b^{\mathrm{T}}=[b_1, \cdots, b_m]$，$A$ 是 $m \times n$ 系数矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

回顾冗长的书写方程组 $AX=b$ 的形式是

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m.$$

系数矩阵 A 的任意非零元可以选作旋转元(选择非零元的原因是，在即将描

述某些旋转步中需要用这个数做除数)，按照下面描述的旋转用这个数得到一个新表. 注意旋转得到的新表所代表的方程组等价于开始的那一个，即初始方程组的任意解是新方程组的一个解，反之亦真.

设 $\alpha=a_{ij}$ 是旋转元(所以 $\alpha\neq 0$)，$x=x_j$ 是与 α 同列(旋转列)的上边缘的元，$u=b_i$ 是与 α 同行(旋转行)的右边缘的元，下面的表描述了这种情况：

$$\begin{array}{ccc} * & x & * \end{array}$$
$$\begin{bmatrix} * & * & * \\ * & \alpha & * \\ * & * & * \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{array}{l} =* \\ =u \\ =* \\ \vdots \end{array}$$

81

特别是，在线性方程组中有了一个形如 $\cdots+\alpha x+\cdots=u$ 的方程. 用这个方程交换表边缘的 x 和 u，对 x 解这个方程. 我们看到，表上边缘 x_i 的个数减少 1，于是离求出关于 x 的方程组的解更近了. 系数矩阵的所有元都将发生某些改变，因而会得到一个新的系数矩阵.

为确定起见，假定 α 在第 1 行、第 1 列(于是 $x=x_1$，$u=u_1$)，注意第 2 行和第 2 列，记 $y=x_2$，$v=b_2$(只有 1 行或 1 列的情形更简单，把它留给读者作为练习)：

$$\begin{array}{ccc} x & y & \cdots \end{array}$$
$$\begin{bmatrix} \alpha & \beta & \cdots \\ \gamma & \delta & \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{array}{l} =u \\ =v \\ \vdots \end{array}$$

首先对 x 解方程 $\alpha x+\beta y+\cdots=u$，得 $x=u/\alpha-\beta y/\alpha-\cdots$，然后将这个对 x 的表达式代入方程组的所有其他方程. 特别地，第 2 个方程 $\gamma x+\delta y+\cdots=v$ 的形式为 $\gamma(u/\alpha-\beta y/\alpha-\cdots)+\delta y+\cdots=v$. 于是得到下面的表：

$$\begin{array}{ccc} u & y & \cdots \end{array}$$
$$\begin{bmatrix} 1/\alpha & -\beta/\alpha & \cdots \\ \gamma/\alpha & \delta-\beta\gamma/\alpha & \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{array}{l} =x \\ =v \\ \vdots \end{array}$$

这证实了前面给出的旋转规则.

从计算上看，像上述这样将表旋转，解线性方程组并不比如第 6 节用增广矩阵更有效. 我们要用旋转表来解线性规划.

在这一节，为了练习表的旋转，我们解一些线性方程组.

例 8.10 考察线性方程组

$$\begin{cases} x+2y=3 \\ 4x+7y=5. \end{cases}$$

将它重写为下面的表：[82]

$$\begin{array}{cc} x & y \end{array}$$
$$\begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} \begin{array}{l} =3 \\ =5 \end{array}$$

选出位于第 1 行第 1 列的元 1 作为旋转元，因为 $1\neq0$，所以这是一个合适的选择，又因为避免了像上面列出的在执行步骤中对分数的处理，所以这也是一个好的选择. 将旋转元用 * 号标记：

$$\begin{array}{cc} x & y \end{array}$$
$$\begin{bmatrix} 1^* & 2 \\ 4 & 7 \end{bmatrix} \begin{array}{l} =3 \\ =5 \end{array}$$

第 1 步　交换边缘的元：

$$\begin{array}{cc} 3 & y \end{array}$$
$$\begin{bmatrix} 1^* & 2 \\ 4 & 7 \end{bmatrix} \begin{array}{l} =x \\ =5 \end{array}$$

第 2 步　旋转元 1 用 $1/1=1$ 代替：

$$\begin{array}{cc} 3 & y \end{array}$$
$$\begin{bmatrix} 1^* & 2 \\ 4 & 7 \end{bmatrix} \begin{array}{l} =x \\ =5 \end{array}$$

第 3 步　旋转行中的元 2 用 $-2/1=-2$ 代替：

$$\begin{array}{cc} 3 & y \end{array}$$
$$\begin{bmatrix} 1^* & -2 \\ 4 & 7 \end{bmatrix} \begin{array}{l} =x \\ =5 \end{array}$$

第 4 步　旋转列的元 4 用 $4/1=4$ 代替：

$$\begin{array}{cc} 3 & y \end{array}$$
$$\begin{bmatrix} 1^* & -2 \\ 4 & 7 \end{bmatrix} \begin{array}{l} =x \\ =5 \end{array}$$

第 5 步　旋转行和旋转列之外仅有的元 7 用 $7-2\cdot4=-1$ 代替：

$$\begin{array}{cc} 3 & y \end{array}$$
$$\begin{bmatrix} 1^* & -2 \\ 4 & -1 \end{bmatrix} \begin{array}{l} =x \\ =5 \end{array}$$

注意新的方程组是

$$\begin{cases} 3-2y=x \\ 4\cdot3-y=5. \end{cases}$$

[83]

(与代入法的各步比较.)

现在选 -1 为旋转元，仍将旋转元用 * 号标记：

$$\begin{array}{cc} 3 & y \end{array}$$
$$\begin{bmatrix} 1 & -2 \\ 4 & -1^{*} \end{bmatrix} \begin{array}{l} = x \\ = 5 \end{array}$$

像上面那样一步一步地旋转，得到下面的表：

$$\begin{array}{cc} 3 & 5 \end{array}$$
$$\begin{bmatrix} -7 & 2 \\ 4 & -1 \end{bmatrix} \begin{array}{l} = x \\ = y \end{array}$$

■

我们作下面两条备注：

(a)矩阵$\begin{bmatrix} -7 & 2 \\ 4 & -1 \end{bmatrix}$是系数矩阵$\begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix}$的逆矩阵，即

$$\begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix}^{-1} = \begin{bmatrix} -7 & 2 \\ 4 & -1 \end{bmatrix}.$$

(b)我们得到了一个上边缘没有变量的表，于是可以以下述方式把两列合成一列，使上边缘的常数为 1：$3 \cdot (-7) + 5 \cdot 2 = -11$；$3 \cdot 4 + 5 \cdot (-1) = 7$. 最后的表给出线性方程组的解：

$$1$$
$$\begin{bmatrix} -11 \\ 7 \end{bmatrix} \begin{array}{l} = x \\ = y \end{array}$$

解为 $x = -11 \cdot 1 = -11$，$y = 7 \cdot 1 = 7$.

这样，我们可以利用旋转来求逆矩阵(若矩阵可逆)并解线性方程组，这可能不是求逆矩阵及解线性方程组的最好的或最有效的方法，但是可以这样做. 下面是利用旋转求解任意线性方程组 $AX = b$ 的更详细的说明. 把方程组写成一个行表

$$X^{\mathrm{T}}$$
$$[A] \quad = b$$

[84]

首先通过反复旋转交换右边缘的常数和上边缘的变量，直到不能再作这样的交换. 然后把所有常数列合到最后一列，其上边缘的常数置 1；如果没有常数列，加一列，上边缘置 1，其余的元置 0. 在考察最后的表之前，我们看看这个常数列如何构成.

假定在交换右边缘的常数和上边缘的变量之后得到下面的表：

$$\begin{array}{ccccc} y_1 & c_1 & y_2 & c_2 & \cdots \end{array}$$
$$\begin{bmatrix} a_{11} & b_{11} & a_{12} & b_{12} & \cdots \\ a_{21} & b_{21} & a_{22} & b_{22} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{array}{l} = z_1 \\ = z_2 \\ \vdots \end{array}$$

其中 y_i 表示变量的行矩阵，c_i 表示常数，a_{ij} 表示矩阵，b_{ij} 表示列矩阵，上边缘为常数 c_i 的列表示常数列，于是在新的线性方程组的第 1 个方程中常数项是

$$c_1 \cdot b_{11} + c_2 \cdot b_{12} + \cdots = b'_1.$$

类似地，第 2 个方程中常数项是

$$c_1 \cdot b_{21} + c_2 \cdot b_{22} + \cdots = b'_2,$$

依次类推. 把这一列写在最后，上边缘的常数置 1，得到最后的表：

$$\begin{matrix} & y & 1 & \\ & \begin{bmatrix} a & b' \\ c & d \end{bmatrix} & & \begin{matrix} = z \\ = e \end{matrix} \end{matrix}$$

这个表中字母 y 和 z 分别表示变量的行矩阵和变量的列矩阵，a 和 c 表示矩阵，d，e 和 b' 表示常数元的列矩阵.

注意矩阵 c 的所有元为零，原因如下：如果 c 有一个非零元，在此处旋转以交换 y 的相应变量和 e 的相应元，但是因为这是最后的表，不可能有这样的交换，因此 c 没有非零元. 还要注意，最后的表中 c 可能不出现(见例 8.10).

如果 $d \neq e$，方程组无解(这是一个不相容方程组). 否则，可以不管最下面这一部分(它得出等式 $d=e$)，方程组的形式为 [85]

$$\begin{matrix} y & 1 & \\ [a & b'] & = z, \end{matrix}$$

即 $z=b'+ay$. 所以 y 中的变量(如果有的话)可取任意值，而 z 中的变量(如果有的话)取相应的值 $z=b'+ay$.

定义 8.11 y 中变量的个数称解集的维数，即解集的维数是矩阵 a 的列数.

这样，任意线性方程组可以用旋转方法求解，能看出旋转的总步数至多是方程组变量的个数吗?

注 8.12 通过用可逆子矩阵 α 代替非零旋转元 α 可以把旋转步推广，仔细地按照正确的次序书写矩阵的乘积，得到下面“推广的旋转方法”：

$$\begin{matrix} x & y \end{matrix} \qquad\qquad\qquad\qquad \begin{matrix} u^{\mathrm{T}} & y \end{matrix}$$
$$\begin{bmatrix} \alpha^* & \beta \\ \gamma & \delta \end{bmatrix} \begin{matrix} = u \\ = v \end{matrix} \quad \mapsto \quad \begin{bmatrix} \alpha^{-1} & -\alpha^{-1}\beta \\ \gamma\alpha^{-1} & \delta-\gamma\alpha^{-1}\beta \end{bmatrix} \begin{matrix} = x^{\mathrm{T}} \\ = v \end{matrix}$$

这一步可以用 k 步原来的旋转和表的行或列的交换来代替，这里 $k \times k$ 是矩阵 α 的大小.

练习

1～2. 利用行表书写问题.

1. 食谱问题(例 2.1).

2. 混合问题(例 2.2).

3. 读下表并将它重写成关于 x'，u，y，z 的线性方程组 $Ax=b$，$x=[x', u, y, z]^{\mathrm{T}}$：

$$\begin{array}{c} \begin{matrix} x' & z & y & 1 \end{matrix} \\ \begin{bmatrix} 3 & 2 & 1 & y \\ -1 & 3 & 0 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{matrix} =u \\ =z \\ =x' \\ =z \end{matrix} \end{array}$$

86

4～7. 在下面的优化问题中所有变量 x，y，$z\geqslant 0$，其余数据由表给出，把问题重写成线性规划的标准型和典范型. 不需要求解优化问题.

4.
$$\begin{array}{c} \begin{matrix} x & y & z & 1 \end{matrix} \\ \begin{bmatrix} 3 & 2 & 1 & 3 \\ -1 & 3 & 0 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{matrix} \geqslant 0 \\ =3 \\ \leqslant x \\ \to \min \end{matrix} \end{array}$$

5.
$$\begin{array}{c} \begin{matrix} x & 3 & y-5z & 1 \end{matrix} \\ \begin{bmatrix} 3 & 2 & 1 & 0 \\ -1 & 3 & 0 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{matrix} \geqslant 3 \\ =y \\ \geqslant 1 \\ \to \min \end{matrix} \end{array}$$

6.
$$\begin{array}{c} \begin{matrix} x & -y & z & -3 \end{matrix} \\ \begin{bmatrix} 3 & 2 & 1 & x \\ -1 & 3 & 0 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{matrix} =0 \\ \geqslant 0 \\ \leqslant 0 \\ \to \max \end{matrix} \end{array}$$

7.
$$\begin{array}{c} \begin{matrix} x & y & z & 1 \end{matrix} \\ \begin{bmatrix} 3 & 2 & 1 & 2 \\ -1 & 3 & 0 & 1 \\ 0 & -2 & 2 & 0 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{matrix} \geqslant 0 \\ \geqslant 0 \\ \geqslant 0 \\ = \to \min \end{matrix} \end{array}$$

8. 在下面表中数值为 1 的元处进行旋转.

$$\begin{array}{c} \begin{matrix} x' & x'' & y & 1 \end{matrix} \\ \begin{bmatrix} 3 & 2 & 1 & b \\ -1 & 3 & 0 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{matrix} =u \\ =a \\ =x' \\ =z \end{matrix} \end{array}$$

(a)数 1 在第 1 行　　　　(b)数 1 在第 1 列

87

9. 用 4 步旋转对练习 1 的矩阵求逆.

10～17. 在下面的每一个表中，对标有 * 号的元进行旋转.

10. $$\begin{array}{c}\begin{matrix}1&a&3&x\end{matrix}\\ \begin{bmatrix}1^*&0&b&a\\-1&2&3&1\end{bmatrix}\begin{matrix}=y\\=z\end{matrix}\end{array}$$

11. $$\begin{array}{c}\begin{matrix}1&a&3&x\end{matrix}\\ \begin{bmatrix}1&0&b&a\\-1^*&2&3&1\end{bmatrix}\begin{matrix}=y\\=z\end{matrix}\end{array}$$

12. $$\begin{array}{c}\begin{matrix}1&a&3&x\end{matrix}\\ \begin{bmatrix}1&0&b&a\\-1&2&3&1^*\end{bmatrix}\begin{matrix}=y\\=z\end{matrix}\end{array}$$

13. $$\begin{array}{c}x\\ [5^*]=2\end{array}$$

14. $$\begin{array}{c}\begin{matrix}1&a&0&x&x\end{matrix}\\ \begin{bmatrix}1&0&b&a&-3\\-1&2^*&3&1&0\end{bmatrix}\begin{matrix}=y\\=z\end{matrix}\end{array}$$

15. $$\begin{array}{c}\begin{matrix}1&0&0&x&1\end{matrix}\\ \begin{bmatrix}1&0&b&a&-3\\-1&2&3&1&0\\-1&2^*&3&1&1\\-1&2&3&1&0\end{bmatrix}\begin{matrix}=y\\=z\\=u\\=v\end{matrix}\end{array}$$

16. $$\begin{array}{c}\begin{matrix}x_1&x_2&x_3&x_4&1\end{matrix}\\ \begin{bmatrix}1&0&1&1&-3\\-1&0&2&1&0\\-1&2&3&1&1\\-1&2&3^*&1&0\\-1&1&0&1&1\end{bmatrix}\begin{matrix}=x_5\\=x_6\\=x_7\\=x_8\\=v\end{matrix}\end{array}$$

17. $$\begin{array}{c}\begin{matrix}x_1&x_2&x_3&x_4&x_5&x_6&1\end{matrix}\\ \begin{bmatrix}1&0&1&1&-3^*&1&0\\-1&0&2&1&0&1&-2\\-1&2&3&1&1&0&0\\-1&2&3&1&0&1&1\\-1&1&0&1&1&2&3\end{bmatrix}\begin{matrix}=x_7\\=x_8\\=x_9\\=x_{10}\\=v\end{matrix}\end{array}$$

18. 设有两个表给出两个方程组

$$\begin{array}{c}y\\ [A]=z\end{array},\quad \begin{array}{c}y\\ [A']=z\end{array},$$

它们在上边缘有同样的变量 y，在右边缘有同样的变量 z，还假设两个方程组是等价的(即有相同的解集)，证明矩阵 A 和 A' 相等.

9. 标准行表

本节用表的形式书写线性规划，然后通过旋转求解. 我们希望找出对于任何线性规划都有效的选择旋转元的方法，也就是当这个方法用于一个线性规划问题时，它能给出最优值和最优解(如果它们存在的话). 简单地说，这个选择旋转元的方法就是，当使用单纯形法这一术语时我们要做的工作.

利用第 7 节的代数技巧，可以约定线性规划中的所有变量非负(目标变量可能除外)，在表中将不再写这些约束，但是不能忘掉它们(把它们写在表的附近).

定义 9.1 标准行表的形式为

$$\begin{array}{cc} x & 1 \end{array}$$
$$\begin{bmatrix} A & b \\ c & d \end{bmatrix} \begin{array}{l} = u \\ = z \mapsto \min, x \geqslant 0, u \geqslant 0, \end{array}$$

其中 A 是系数矩阵，b 是表示线性约束右端项的列矩阵，c 是表示目标函数系数的行矩阵，d 是一个给定的数，x 是变量的行矩阵，u 是变量的列矩阵，z(可写可不写)是目标变量，x，u，z 中的所有变量是不同的.

更准确些，一个表是标准的，如果

- 系数矩阵不含任何未知数(只有给定的数或者有参数).
- 每一行表示一个线性方程(不是不等式)，最后一行可能例外.
- 最后一行表示要最小化的目标函数.
- 表边缘的所有变量(目标变量可能例外)非负(将把这些非负约束写在表的附近，但不是表的一部分).
- 表边缘的所有变量都是不同的独立变量.
- 表上边缘刚好有一个常数，它是 1，在最后一列上.

89

如定义 9.1 所述，目标变量在标准表中不一定写出.

这样，在明确地陈述问题和收集数据之后，用单纯形法求解线性规划包括以下步骤：

1. 将问题写成标准表的形式(如果需要，改换变量使表中只含非负变量，目标变量可能除外).
2. 利用旋转直到得到最优解或者发现无解(如何选择旋转元留待以后讨论).
3. 用原来变量写出解答.
4. 用解答对原来的问题做出解释. 如果发现解答不满意，要调整模型，收集更多数据，解相应的线性规划问题，等等.

例 9.2 设以 x，y，z，w 为变量的线性规划由下表给出(没有其他任何约束)：

$$
\begin{array}{cccc} 2 & -x & y & -1 \end{array}
$$

$$
\begin{bmatrix} z & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 \\ z & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{array}{l} = w \to \max \\ \geqslant 0 \\ \geqslant 1 \\ = y \end{array}
$$

这个表不是标准的，事实上，我们刚才讨论的每一条规定都被破坏，即

- 没有约束 $x \geqslant 0$.
- 变量 z 在矩阵中(标准表中它在边缘).
- 表的第 2 行和第 3 行有$\geqslant$符号.
- 目标函数出现在矩阵的第 1 行，而不是最后一行.
- 是最大化问题而不是最小化问题.
- 变量 y 出现两次，变量 x 带系数-1 而不是 1 出现. [90]
- 边缘有太多的常数.
- 矩阵最后一列上面的常数是-1 而不是 1.

如何把这个表整理成标准型呢? ■

实际上，有很多方法利用标准行表重写这个线性规划问题. 例如，读这个表得到

$$
\begin{cases} 2z + y - 2 = -w \to \max \\ y \geqslant 0 \\ 2z \geqslant 1 \\ 2 - 2x + 3y - 4 = y, \end{cases}
$$

或

$$
\begin{cases} -2z - y + 2 = -w \to \min \\ y \geqslant 0 \\ z \geqslant 1/2 \\ x = y - 1. \end{cases}
$$

引入松弛变量 u，用 $z - 1/2 = u \geqslant 0$ 代替 $z \geqslant 1/2$，将问题写成标准型：

$$
\begin{array}{ccc} y & z & 1 \end{array}
$$

$$
\begin{bmatrix} 0 & 1 & -1/2 \\ -1 & -2 & 2 \end{bmatrix} \begin{array}{l} = u \\ = w \to \min \end{array}
$$

且 $y \geqslant 0$，$z \geqslant 0$，$u \geqslant 0$，$x = y - 1$.

注意，这个表不包含变量 x，且附加信息($\to \min$，$y \geqslant 0$，$z \geqslant 0$，$u \geqslant 0$，$x = y - 1$)写在表外.

例 9.3 设有线性规划问题如下：

$$\begin{array}{cccccc} v & 1 & x & y & x & -2 \end{array}$$

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & -7 & 8 & -9 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 & 2 & 0 \\ 0 & 3 & 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \begin{array}{l} =0 \\ \geqslant 0 \\ \geqslant 0 \\ \geqslant 0 \\ \leqslant 0 \\ =z \to \max \\ =u \\ =u \end{array}$$

91

我们将这个表变成标准型. 首先合并常数列(即上边缘标有 1 和 -2 的列)，得到列

$$1 \cdot 1 + (-2 \cdot 5) = -9$$
$$1 \cdot 0 + (-2 \cdot 1) = -2$$
$$1 \cdot 0 + (-2 \cdot 0) = 0$$
$$1 \cdot 3 + (-2 \cdot 0) = 3$$
$$1 \cdot 3 + (-2 \cdot 0) = 3$$
$$1 \cdot 0 + (-2 \cdot 0) = 0$$
$$1 \cdot 0 + (-2 \cdot 0) = 0$$
$$1 \cdot 0 + (-2 \cdot 0) = 0.$$

将这个常数列写在表的最后一列，上边缘标以系数 1.

$$\begin{array}{ccccc} v & x & y & x & 1 \end{array}$$

$$\begin{bmatrix} 0 & 2 & 3 & 4 & -9 \\ 0 & -7 & 8 & -9 & -2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 3 \\ 0 & 1 & 0 & 2 & 3 \\ 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & -1 & 0 \end{bmatrix} \begin{array}{l} =0 \\ \geqslant 0 \\ \geqslant 0 \\ \geqslant 0 \\ \leqslant 0 \\ =z \to \max \\ =u \\ =u \end{array}$$

然后合并标有 x 的两列为一列，系数矩阵的这一列由两列对应的元相加得到

$$2 + 4 = 6$$
$$-7 - 9 = -16$$
$$0 + 0 = 0$$
$$1 + 2 = 3$$

$$1+2=3$$
$$0+1=1$$
$$1+2=3$$
$$0-1=-1.$$

92

于是得到下面的表：

$$\begin{array}{c} \begin{matrix} v & y & x & 1 \end{matrix} \\ \begin{bmatrix} 0 & 3 & 6 & -9 \\ 0 & 8 & -16 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 3 & 3 \\ 1 & 2 & 1 & 0 \\ 0 & 0 & 3 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{matrix} =0 \\ \geqslant 0 \\ \geqslant 0 \\ \geqslant 0 \\ \leqslant 0 \\ =z\to \max \\ =u \\ =u \end{matrix} \end{array}$$

再把目标函数的系数写到矩阵的最后一行：

$$\begin{array}{c} \begin{matrix} v & y & x & 1 \end{matrix} \\ \begin{bmatrix} 0 & 3 & 6 & -9 \\ 0 & 8 & -16 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 3 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 2 & 1 & 0 \end{bmatrix} \begin{matrix} =0 \\ \geqslant 0 \\ \geqslant 0 \\ \geqslant 0 \\ \leqslant 0 \\ =u \\ =u \\ =z\to \max \end{matrix} \end{array}$$

还可以把两个 u 行减少一个，得到这样两行，一行是原来在右边缘标以 u 的，另一行要在右边缘标以 0. 这样做是因为要在两行边缘有不同的变量：

$$\begin{array}{c} \begin{matrix} v & y & x & 1 \end{matrix} \\ \begin{bmatrix} 0 & 3 & 6 & -9 \\ 0 & 8 & -16 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 3 & 0 \\ 1 & 0 & -4 & 0 \\ 1 & 2 & 1 & 0 \end{bmatrix} \begin{matrix} =0 \\ \geqslant 0 \\ \geqslant 0 \\ \geqslant 0 \\ \leqslant 0 \\ =u \\ =0 \\ =z\to \max \end{matrix} \end{array}$$

因为不想在右边缘有常数项，对上面矩阵第 2 行和第 4 行的不等式引入新的非负(松弛或剩余)变量 w_1 和 w_2. 注意第 5 行的不等式为 $3x+3\leqslant 0$，用等价的不等式 $-3x-3\geqslant 0$ 代替它，然后引入非负松弛变量 w_3，于是得等式 $-3x-3=w_3$. 因为第 3 行 $y\geqslant 0$ 正好是对变量 y 的非负约束，把它分开写在表的旁边. 再交换标以 x 和 y 的两列，新的表为

[93]

$$\begin{array}{cccc} v & x & y & 1 \\ \end{array}$$
$$\begin{bmatrix} 0 & 6 & 3 & -9 \\ 0 & -16 & 8 & -2 \\ 0 & 3 & 0 & 3 \\ 0 & -3 & 0 & -3 \\ 0 & 3 & 0 & 0 \\ 1 & -4 & 0 & 0 \\ 1 & 1 & 2 & 0 \end{bmatrix} \begin{array}{l} =0 \\ =w_1 \\ =w_2 \\ =w_3 \\ =u \\ =0 \\ =z\to\max \end{array}$$

且 $y\geqslant 0$，$w_1\geqslant 0$，$w_2\geqslant 0$，$w_3\geqslant 0$.

现在把约束 $u=3x$ 拿到表外，分开写在表的旁边，并进行行的交换使右边缘标以常数的行先出现：

$$\begin{array}{cccc} v & x & y & 1 \\ \end{array}$$
$$\begin{bmatrix} 0 & 6 & 3 & -9 \\ 1 & -4 & 0 & 0 \\ 0 & -16 & 8 & -2 \\ 0 & 3 & 0 & 3 \\ 0 & -3 & 0 & -3 \\ 1 & 1 & 2 & 0 \end{bmatrix} \begin{array}{l} =0 \\ =0 \\ =w_1 \\ =w_2 \\ =w_3 \\ =z\to\max \end{array}$$

且 $y\geqslant 0$，$w_1\geqslant 0$，$w_2\geqslant 0$，$w_3\geqslant 0$，$u=3x$.

右边缘仍有两个常数，在第 2 行的 1 和第 1 行的 6 处进行旋转，这些常数就会出现在上边缘：

$$\begin{array}{cccc} 0 & 0 & y & 1 \\ \end{array}$$
$$\begin{bmatrix} 0 & 1/6 & -1/2 & 3/2 \\ 1 & 2/3 & -2 & 6 \\ 0 & -8/3 & 16 & -26 \\ 0 & 1/2 & -3/2 & 15/2 \\ 0 & -1/2 & 3/2 & -15/2 \\ 1 & 5/6 & -1/2 & 15/2 \end{bmatrix} \begin{array}{l} =x \\ =v \\ =w_1 \\ =w_2 \\ =w_3 \\ =z\to\max \end{array}$$

[94]

且 $y\geqslant 0$，$w_1\geqslant 0$，$w_2\geqslant 0$，$w_3\geqslant 0$，$u=3x$.

为了将最大化问题转化为最小化问题，最后一行乘 -1：

$$\begin{array}{cccc|l} 0 & 0 & y & 1 & \\ \hline \end{array}$$

$$\begin{array}{cccc} 0 & 0 & y & 1 \end{array}$$
$$\begin{bmatrix} 0 & 1/6 & -1/2 & 3/2 \\ 1 & 2/3 & -2 & 6 \\ 0 & -8/3 & 16 & -26 \\ 0 & 1/2 & -3/2 & 15/2 \\ 0 & -1/2 & 3/2 & -15/2 \\ -1 & -5/6 & 1/2 & -15/2 \end{bmatrix} \begin{array}{l} =x \\ =v \\ =w_1 \\ =w_2 \\ =w_3 \\ =-z\to\min \end{array}$$

且 $y\geqslant 0$，$w_1\geqslant 0$，$w_2\geqslant 0$，$w_3\geqslant 0$，$u=3x$.

变量 x 和 v 不要求非负，所以它们不能出现在标准表中. 另一方面，对于这些变量没有任何限制，可以把前两个约束放在表外. 上边缘标以 0 的列可以去掉. 下面是表示线性规划的标准表：

$$\begin{array}{cc} y & 1 \end{array}$$
$$\begin{bmatrix} 16 & -26 \\ -3/2 & 15/2 \\ 3/2 & -15/2 \\ 1/2 & -15/2 \end{bmatrix} \begin{array}{l} =w_1 \\ =w_2 \\ =w_3 \\ =-z\to\min, \end{array}$$

且 $y\geqslant 0$，$w_1\geqslant 0$，$w_2\geqslant 0$，$w_3\geqslant 0$，$u=3x$，$x=3/2-y/2$，$v=6-2y$. ■

从一般线性规划典范型到标准表及其反变换

通常将一个线性规划写成标准表比较容易，例如当线性规划以典范型

$$\text{minimize } cx+d, \text{subject to } Ax\leqslant b, x\geqslant 0,$$

给出时，其步骤如下：

1. 将约束 $Ax\leqslant b$ 换成等价约束 $-Ax\geqslant -b$.

2. 约束 $-Ax\geqslant -b$ 可看作约束 $-Ax+b\geqslant 0$，为了得到等式而非不等式，引入松弛变量 u，定义为 $u=-Ax+b$. 注意现在要求 $x\geqslant 0$，$u\geqslant 0$. 标准行表为

95

$$\begin{array}{cc} x^{\mathrm{T}} & 1 \end{array}$$
$$\begin{bmatrix} -A & b \\ c & d \end{bmatrix} \begin{array}{l} =u \\ \to\min, \quad x\geqslant 0, \quad y\geqslant 0 \end{array}$$

其中 u 是松弛变量的列矩阵（u 中的所有变量不同，且与列矩阵 x 中的变量也不同）.

从一个标准表到线性规划典范型也很容易（去掉右边缘的变量），所以标准表与典范型之间几乎没有差别.

从一般线性规划标准型到标准表及其反变换

如果线性规划以标准型

$$\text{minimize } cx+d, \text{subject to } Ax=b, x\geqslant 0$$

给出，则至少有两种办法把它表示成标准表：

（ⅰ）（人工变量）将问题变成典范型

$$\text{minimize } cx+d, \text{subject to } Ax\leqslant b, -Ax\leqslant -b, x\geqslant 0,$$

做法同上.

由此得到标准表

$$\begin{matrix} x^{\mathrm{T}} & 1 \end{matrix}$$
$$\begin{bmatrix} -A & b \\ A & -b \\ c & d \end{bmatrix} \begin{matrix} =u \\ =v \\ \to \min, x\geqslant 0, u\geqslant 0, v\geqslant 0 \end{matrix}$$

列 u，v 中的变量称为人工变量，它们与技巧 7.7 中的人工变量根据对偶关系相联系，见第 4 章.

（ⅱ）解方程组 $Ax=b$（这一步可以用前面描述的的旋转方法来做），或者发现方程组无解，于是线性规划不可行；或者得到形如 $y=Dz+e$ 的解，其中 y 是 x 中的某些变量，z 是 x 中其余的变量，然后将目标函数用 z 表示为 $cx+d=c'z+$ 96 d'，得到标准行表

$$\begin{matrix} z^{\mathrm{T}} & 1 \end{matrix}$$
$$\begin{bmatrix} D & e \\ c' & d' \end{bmatrix} \begin{matrix} =y \\ \to \min, z\geqslant 0, y\geqslant 0 \end{matrix}$$

从任意标准表变回到相应的线性规划标准型更容易：只需把右边缘的变量放到相应方程的左边，把常数项放到这些方程的右边，就是说，将定义 9.1 中的 $Ax+b=u$ 重写作 $Ax-u=-b$.

如果一个线性规划有某些变量不要求非负，那么可以从中消去它们得到标准型.

例 9.4 将以下问题用标准表写出：

$$x+y+z\to \min, 2x+3y+4z\leqslant 5, x-y+4z=3, x\geqslant 0, y\geqslant 0.$$

解 求解关于 z 的方程，将它从问题中消去：

$$z=(3-x+y)/4,$$

$$x+y+z=x+y+(3-x+y)/4=3x/4+5y/4+3/4\to \min,$$

$$2x+3y+4z=2x+3y+4(3-x+y)/4=x+4y+3\leqslant 5, x\geqslant 0, y\geqslant 0,$$

或

$$3x/4+5y/4+3/4\to \min, x+4y\leqslant 2, x\geqslant 0, y\geqslant 0.$$

带两个变量 x，y 的问题以典范型给出，可以用标准表写为

$$\begin{matrix} x & y & 1 \end{matrix}$$
$$\begin{bmatrix} -1 & -4 & 2 \\ 3/4 & 5/4 & 3/4 \end{bmatrix} \begin{matrix} =u \\ \to \min \end{matrix}, x\geqslant 0, y\geqslant 0, z=(3-x+y)/4.$$

将 $z=(3-x+y)/4$ 放在边上，z 的值可包含在最终的解中. 显然这个优化问题

恰有一个最优解：$x=0$，$y=0$，$\min=3/4$，$z=3/4$（虽然最优解的唯一性不是问题所要求的）. ■

例 9.5 将以下问题用标准表写出：

$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & 3 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 1 & -3 & 1 & 0 \\ -1 & 0 & 2 & 1 & 0 & 1 & -2 \\ -1 & 2 & 3 & 1 & 1 & 0 & 0 \\ -1 & 2 & 3 & 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 & 1 & 2 & 3 \end{bmatrix} \begin{array}{l} = x_7 \\ = x_8 \\ = x_9 \\ = x_{10} \\ = v \to \max \end{array},$$

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_{10} \geqslant 0.$$

解 从表中消去 x_8，x_9，目标变量 v 乘以 -1，最后一列乘以 3，得

$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & 1 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 1 & -3 & 1 & 0 \\ -1 & 2 & 3 & 1 & 0 & 1 & 3 \\ 1 & -1 & 0 & -1 & -1 & -2 & -9 \end{bmatrix} \begin{array}{l} = x_7 \\ = x_{10} \\ = -v \to \min \end{array},$$

$$x_1, x_2, x_3, x_4, x_5, x_6 \geqslant 0,$$

$$x_8 = -x_1 + 2x_3 + x_4 + x_6 - 6, x_9 = -x_1 + 2x_2 + 3x_3 + x_4 + x_5.$$ ■

练习

1. 解例 9.3 的线性规划，用原变量 x，u 和 v 给出解.

2～10. 将下面的优化问题用标准行表重写成线性规划. 不需要解优化问题. 提示：首先把问题写成线性规划典范型.

2. $$\begin{cases} \text{maximize} & P=2x+3y \\ \text{subject to} & x\geqslant 0,\ y\geqslant 0 \\ & 4x+5y\leqslant 7. \end{cases}$$

3. $$\begin{cases} \text{maximize} & x \\ \text{subject to} & |x+y|\leqslant 1 \\ & |x-y|\leqslant 1. \end{cases}$$

4. $$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & -1 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 1 & -3 & 1 & 0 \\ -1 & 0 & 2 & 1 & 0 & 1 & -2 \\ -1 & 2 & 3 & 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 1 & 2 & 3 \end{bmatrix} \begin{array}{l} \geqslant 0 \\ = x_8 \\ = 0 \\ = v \to \min \end{array}$$
$$x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7,\ x_8,\ x_9,\ x_{10} \geqslant 0.$$

5.
$$\begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & 1 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 1 & -3 & 1 & 0 \\ -1 & 0 & 2 & 1 & 0 & 1 & -2 \\ -1 & 2 & 3 & 1 & 1 & 0 & 0 \\ 0 & 2 & 3 & 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 & 1 & 2 & 3 \end{bmatrix} \begin{matrix} =x_7 \\ =x_8 \\ =x_9 \\ =x_2 \\ =v\to\max \end{matrix}$$
$x_1, x_2, x_3, x_4, x_5, x_6, x_7 \geqslant 0.$

6.
$$\begin{array}{ccccccc} 2x_1 & x_2 & -3x_1 & x_4 & x_5 & x_6 & 1 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 1 & -3 & 1 & 0 \\ -1 & 0 & 2 & 1 & 0 & 1 & -2 \\ 0 & 2 & 3 & 1 & 1 & 0 & 0 \\ -1 & 2 & 3 & 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 & 1 & 2 & 3 \end{bmatrix} \begin{matrix} =x_7 \\ =x_7 \\ \leqslant 0 \\ =x_3 \\ =v\to\min \end{matrix}$$
$x_1, x_2, x_3, x_4, x_5, x_6, x_7 \geqslant 0.$

7.
$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & 1 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 1 & -3 & 1 & 0 \\ -1 & 0 & 2 & 1 & 0 & 1 & -2 \\ -1 & 2 & 3 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} =x_2 \\ =x_4 \\ =v\to\min \end{matrix}$$
$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10} \geqslant 0.$

8.
$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & -4 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 1 & -3 & 1 & 0 \\ -1 & 0 & 2 & 1 & 0 & 1 & -2 \\ -1 & 2 & 3 & 1 & 1 & 0 & 0 \\ -1 & 2 & 3 & 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 & 1 & 2 & 3 \end{bmatrix} \begin{matrix} \geqslant 0 \\ =1 \\ =0 \\ =x_1 \\ =v\to\min \end{matrix}$$
$x_1, x_2, x_3, x_4, x_5, x_6 \geqslant 0.$

9.
$$\begin{array}{ccccccc} x_7 & x_2 & x_3 & x_4 & x_5 & x_6 & 1 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 1 & -3 & 1 & 0 \\ -1 & 0 & 2 & 1 & 0 & 1 & -2 \\ -1 & 2 & 3 & 1 & 1 & 0 & 0 \\ -1 & 2 & 3 & 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 & 1 & 2 & 3 \end{bmatrix} \begin{matrix} =0 \\ =0 \\ =0 \\ =x_1 \\ =v\to\min \end{matrix}$$
$x_1, x_2, x_3, x_4, x_5, x_6 \geqslant 0.$

10.
$$
\begin{array}{ccccccc}
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & -1 \\
\end{array}
$$
$$
\begin{bmatrix}
1 & 0 & 1 & 1 & -3 & 1 & 0 \\
-1 & 0 & 2 & 1 & 0 & 1 & -2 \\
-1 & 2 & 3 & 1 & 1 & 0 & 0 \\
-1 & 2 & 3 & 1 & 0 & 1 & 1 \\
-1 & 1 & 0 & 1 & 1 & 2 & 3
\end{bmatrix}
\begin{array}{l}
=x_1 \\
=2x_1 \\
=x_7 \\
=x_2 \\
=v\rightarrow \max
\end{array}
$$
$x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6 \geqslant 0.$

99

第4章　单纯形法

10. 单纯形法的第 2 阶段

用标准行表表示的单纯形法，可以看作是为了经过有限次旋转达到终止表而选取旋转元的一种方法．通常，单纯形法分为两个阶段．第 1 阶段的目的是找一个可行解；第 2 阶段的目的是设法改进目标函数，并保持可行性直到找到最优解．本节我们将详细地描述第 2 阶段．第 1 阶段与第 2 阶段类似，但要复杂些，我们将在下一节详细描述．我们可以运用一些技巧，将第 1 阶段的问题转化为第 2 阶段的问题，也可以将第 2 阶段的问题转化为第 1 阶段的问题．对偶的观点表明这两个阶段是可以相互转化的(参见第 5 章).

考虑以下写成标准行表的线性规划：

$$\begin{array}{cc} x & 1 \end{array}$$
$$\begin{bmatrix} A & b \\ c & d \end{bmatrix} \begin{array}{l} = u \\ = z \to \min, x \geqslant 0, u \geqslant 0 \end{array} \tag{10.1}$$

其中 A 是系数矩阵，b 是由线性约束的右端项组成的列向量，c 是由目标函数系数组成的行向量，d 是一个给定的数，x 是由变量组成的行向量(称为**非基变量**，nonbasic variable)，u 是由变量组成的列向量(称为**基变量**，basic variable)，z 是目标变量，x，u，z 的所有分量都是不同的变量．

这个线性规划写成矩阵形式就是

$$\text{minimize } z = cx^{\mathrm{T}} + d, \text{subject to } Ax^{\mathrm{T}} + b = u \geqslant 0, x \geqslant 0.$$

如果由线性规划的典范型

$$\text{minimize } z = cx^{\mathrm{T}} + d, \text{subject to} - Ax^{\mathrm{T}} \geqslant b,\ x \geqslant 0$$

得到它的标准表，那么 u 是松弛变量．但是，经过一些旋转步以后，有的松弛变量可能会变成非基变量(即它们能转变成表的最上边的变量)．此时，我们称这些变量以及在标准型里对应的不等式约束是**有效的**(active).　[100]

注　一般地，在优化问题中，一个可行解，若使一个不等式约束的等号成立，那么这个不等式约束称为有效的(active)或紧的(tight)．如果一个约束去掉后，原最优区域改变(在更严格的意义下要求最优值发生改变)，则称该约束为绑定(binding)约束．就线性规划问题的最优解来说，任何绑定约束都是紧的．

注意线性方程组 $A^{\mathrm{T}} \cdot x + b = u$ 有解 $x = 0$，$u = b$．对给定的标准行表而言，此解称为**基解**(basic solution)，对应的目标函数值是 d．这说明下面的定义是合理的：

定义 10.2　若 $b \geqslant 0$（即基解是可行的），则称标准表（见(10.1)）是行可行的(row feasible).

例如，下面两个标准表都是可行的：

$$\begin{array}{cc} x & 1 \\ \end{array}\qquad\qquad 1$$
$$\begin{bmatrix} -1 & 1 \\ 0 & 0 \\ -2 & -3 \end{bmatrix} \begin{array}{l} = u \\ = v \\ \to \min \end{array}, \qquad \begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix} \begin{array}{l} = x \\ = y \\ \to \min \end{array} \qquad (10.3)$$

对应的基解是

$$x = 0, u = 1, v = 0 \text{ 和 } x = 2, y = 3.$$

它们都是可行的．下面的两个标准表都是不可行的：

$$\begin{array}{cc} x & 1 \\ \end{array}\qquad\qquad 1$$
$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \\ -2 & -3 \end{bmatrix} \begin{array}{l} = u \\ = v \\ \to \min \end{array}, \qquad \begin{bmatrix} -2 \\ 3 \\ -2 \end{bmatrix} \begin{array}{l} = x \\ = y \\ \to \min \end{array} \qquad (10.4)$$

对应的基解是

$$x = 0, u = 1, v = -1 \text{ 和 } x = -2, y = 3.$$

这些解都是不可行的．

101

定义 10.5　若 $b \geqslant 0$ 且 $c \geqslant 0$，前面的表，即式(10.1)，称为最优的．

此时，基解就是最优解，d 是最优值．事实上，一方面基解 $x=0$，$u=b$ 是可行的(因为 $b \geqslant 0$)且目标函数 z 在此基解上的值是 d. 另一方面，因为 $c \geqslant 0$，故对任意的行向量 $x \geqslant 0$，$z=c^{\mathrm{T}}x+d \geqslant d$ 均成立，因此对 z 我们不能得到更小的值了(即 d 是 z 能达到的最小值). 例如式(10.3)中第一个表不是最优的，第二个是最优的，且 $\min = -2$.

事实上，最优表是通过一个线性方程组来描述线性规划的所有最优解．在任一最优解中，与 c 的非零元对应的非基变量必须是零．对其他变量来说，表中所给出的线性方程组加上这些变量的非负限制条件，是保证最优性的充分必要条件. 而且，任何一个最优解都包含基最优解的其他有用的信息(参见第 5 章).

于是，如果一个线性规划可由一个最优表来描述，那么它是可解的．单纯形法试图由任意(标准)表经有限次旋转得到一个最优表．旋转元不能从最后列(在最上边标有常数“1”的列)和最后行(代表目标函数的行)中选取，因此旋转步将一个标准表变成另一个标准表．

单纯形法的第 2 阶段开始于一个可行表，最后要么产生一个最优表，要么产生一个具有坏列(bad column)的表，这里坏列的定义如下．

定义 10.6　标准表中的一个列，如果它不是最后列(在顶端标注 1)且最后一个元素是负的，而其他元素均是非负的，则该列称为坏列．

例如式(10.3)和式(10.4)中的表都没有坏列，但下面的两个表都有一个坏列：

$$\begin{array}{cc} y & 1 \end{array}\qquad\qquad \begin{array}{ccc} x & y & 1 \end{array}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ -3 & -1 \end{bmatrix} \begin{matrix} =u \\ =v \\ =\to \min \end{matrix},\qquad \begin{bmatrix} 0 & -1 & 1 \\ 0 & 1 & -3 \\ -1 & -3 & -1 \end{bmatrix} \begin{matrix} =u \\ =v \\ =\to \min \end{matrix}$$

如果表中有一个坏列，可以让它顶端的变量取越来越大的值，顶端其他的变量(如果存在的话)设为0. 这样可以得到使目标函数值越来越小的可行解．此时问题无界，没有最优解．特别是不能得到一个最优表．“问题无界”简记为 $\min=-\infty$.

102

本节我们详细描述单纯形法的第2阶段，并证明它总在有限步终止．

单纯形法是恰当选取旋转元并经过有限次旋转达到终止表的一种方法．特别地，在第2阶段，所有的表格都是标准的和可行的，每次旋转后目标函数值(在基解上)得到改进或保持不变(在退化情况下)，经过有限次旋转得到的终止表要么是最优的，要么有一个坏列．

我们看可行的标准表，即式(10.1)，由定义10.2知可行性是指 $b\geqslant 0$. 我们希望通过旋转得到一个最优表．单纯形法包括检验最终表(求解结束)，选取旋转元和旋转过程．可以把一个旋转步包含的过程分成四小步．

单纯形法的第2阶段

1. 若 $c\geqslant 0$，则可行表已经是最优的．此时，$\min=d$，$x=0$，$u=b$. 结束！
2. 检验是否有坏列．即对每个 $c_j<0$，检查矩阵 A 的第 j 列的所有元素是否非负．如果找到一个坏列，我们可得答案：$\min=-\infty$，$x_j\to\infty$，令顶端的其他变量(非基变量) $x_i=0$，$i\neq j$，且 $u=A\cdot x+b$. 结束！
3. 取某一个 $c_j<0$，考虑第 j 列(将作为旋转列)中所有取值为负的元素 $a_{ij}<0$. 对这些 a_{ij}，选取 b_i/a_{ij} 中最大者(例如比值最接近0)，记作 $a_{i'j}$，作为旋转元．
4. 进行旋转，转到1.

103

通过旋转，我们可以得到一个新的可行的标准表．事实上，$b_{i'}\to -b_{i'}/a_{i'j}\geqslant 0$. 若 $i\neq i'$，则有 $b_i\mapsto \bar{b}_i=b_i-a_{ij}\cdot b_{i'}/a_{i'j}$. 若 $a_{ij}\geqslant 0$，则 $\bar{b}_i\geqslant b_i\geqslant 0$. 若 $a_{ij}<0$，则有 $\bar{b}_i\geqslant 0$，因为 $b_i/a_{ij}\leqslant b_{i'}/a_{i'j}$. 目标函数 z 的值得到改进或保持不变：它由 d 变成 $d-c_jb_{i'}/a_{i'j}\leqslant d$，其中 $a_{i'j}$ 是旋转元．它保持不变当且仅当 b_i 的最后一个元素在旋转行中为0. 此时旋转步是退化的(degenerate). 一个退化的旋转步不会改变基可行解但能改变基(在表中最右端的变量集合).

问题 10.7 求解下列由标准行表给出的线性规划：

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & 1 \end{array}$$

(a) $$\begin{bmatrix} 1 & -1 & 0 & 1 & 0 \\ -1 & 1 & -2 & -1 & -2 \\ 0 & 1 & 2 & -1 & -2 \end{bmatrix} \begin{matrix} =e \\ =f \\ \to\min \end{matrix}$$

(b)
$$\begin{array}{ccccc|l} x_1 & x_2 & x_3 & x_4 & 1 & \\ 1 & -1 & 0 & 1 & 0 & =e \\ -1 & 1 & -2 & -1 & 1 & =f \\ 1 & 1 & -2 & 0 & -2 & \to\min \end{array}$$

(c)
$$\begin{array}{ccccc|l} x_1 & x_2 & x_3 & x_4 & 1 & \\ 1 & 1 & 0 & 1 & 0 & =e \\ -1 & 1 & -2 & -1 & 1 & =f \\ 0 & -1 & 2 & 0 & 2 & \to\min \end{array}$$

(d)
$$\begin{array}{ccccc|l} x_1 & x_2 & x_3 & x_4 & 1 & \\ 1 & 1 & 0 & 1 & 0 & =e \\ -1 & -1 & -2 & -1 & 1 & =f \\ 0 & -1 & 2 & 0 & 2 & \to\min \end{array}$$

解

(a)此表不可行，不能用第 2 阶段求解．我们将在后面解决此类问题．

(b)此表是最优的，因此有 min＝－2，在 $x_1=x_2=x_3=x_4=0$，$e=0$，$f=1$ 达到．如果需要，可以给出所有最优解：$x_1=x_2=x_3=0$，$0\leqslant x_4\leqslant 1$，$e=x_4$，$f=1-x_4$．

104

(c)1. 此表不是最优的；2. x_2 所在列是坏列，故该问题无界(min＝－∞)．

(d)1. 此表不是最优的；2. 没有坏列；3. 旋转列是 x_2 所在列，旋转元是 f 所在行的－1.

$$\begin{array}{ccccc|l} x_1 & x_2 & x_3 & x_4 & 1 & \\ 1 & 1 & 0 & 1 & 0 & =e \\ -1 & -1^* & -2 & -1 & 1 & =f \\ 0 & -1 & 2 & 0 & 2 & \to\min \end{array}$$

4. 首先计算最后一行和最后一列的新元素：

$$\begin{array}{ccccc|l} x_1 & f & x_3 & x_4 & 1 & \\ & & & & 1 & =e \\ & & & & 1 & =x_2 \\ 1 & 1 & 4 & 1 & 1 & \to\min \end{array}$$

可以看到新得到的表是最优的，min＝1 在 $x_1=f=x_3=x_4=0$，$e=1$，$x_2=1$ 达到．注意这里旋转保持了表是标准可行的，并且改变了目标函数值，这是第 2 阶段应做到的．

■

问题 10.8 求解

$$\begin{array}{ccccc} x_1 & -x_2 & 2x_3 & x_4 & -2 \end{array}$$
$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ -1 & -1 & -2 & -1 & 1 \\ 0 & -1 & 2 & 0 & 2 \end{bmatrix} \begin{array}{l} = 3x_5 \\ = -x_6 \\ \to \min \end{array} \qquad \text{所有 } x_i \geqslant 0$$

解 此表不是标准型．为了将其化为标准型，将第 2 列乘 -1，最后一列乘 -2，第 2 行乘 -1，得

$$\begin{array}{ccccc} x_1 & x_2 & 2x_3 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} 1 & -1 & 0 & 1 & 0 \\ 1 & -1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 0 & -4 \end{bmatrix} \begin{array}{l} = 3x_5 \\ = x_6 \\ \to \min \end{array} \qquad \text{所有 } x_i \geqslant 0$$

不必对第 3 列和第 1 行进行归一化处理，也不必用别的变量替换 $2x_3$ 和 $3x_5$．新表是最优的，可直接得答案：$x_1=x_2=x_3=x_4=0$，$x_5=0$，$x_6=2$，$\min=-4$．

■

问题 10.9 求解

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ -1 & -1 & -2 & -1 & 1 \\ 0 & -1 & 2 & 0 & 2 \end{bmatrix} \begin{array}{l} = x_5 \\ = 0 \\ \to \min \end{array} \qquad \text{所有 } x_i \geqslant 0$$

105

解 此表不是标准型．为了将其化为标准型，选取第 2 行的第一个元素 -1 做旋转元．旋转将 x_1 和 0 进行了交换，仅有矩阵中的第 1 行改变了：

$$\begin{array}{ccccc} 0 & x_2 & x_3 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} -1 & 0 & -2 & 0 & 1 \\ -1 & -1 & -2 & -1 & 1 \\ 0 & -1 & 2 & 0 & 2 \end{bmatrix} \begin{array}{l} = x_5 \\ = x_1 \\ \to \min \end{array} \qquad \text{所有 } x_i \geqslant 0$$

去掉第 1 列可得标准型

$$\begin{array}{cccc} x_2 & x_3 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} 0 & -2 & 0 & 1 \\ -1 & -2 & -1 & 1 \\ -1 & 2 & 0 & 2 \end{bmatrix} \begin{array}{l} = x_5 \\ = x_1 \\ \to \min \end{array} \qquad \text{所有 } x_i \geqslant 0$$

因为此表是标准的，可使用单纯形法的第 2 阶段．此时只有一个选择：交换 x_1 和 x_2，得

$$\begin{array}{cccc} x_2 & x_3 & x_4 & 1 \end{array} \qquad\qquad \begin{array}{cccc} x_1 & x_3 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} 0 & -2 & 0 & 1 \\ -1^* & -2 & -1 & 1 \\ -1 & 2 & 0 & 2 \end{bmatrix} \begin{array}{l} = x_5 \\ = x_1 \\ \to \min \end{array} \quad \mapsto \quad \begin{bmatrix} 0 & -2 & 0 & 1 \\ -1 & -2 & -1 & 1 \\ 1 & 4 & 1 & 1 \end{bmatrix} \begin{array}{l} = x_5 \\ = x_2 \\ \to \min \end{array}$$

此表是最优的，可得答案：$x_1=x_3=x_4=0$，$x_2=1$，$x_5=1$，$\min=1$. 还有其他的最优解，但不必写出． ■

问题 10.10 求解

$$\begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & 1 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ -1 & -1 & -2 & -1 & 1 \\ 0 & -1 & 2 & 0 & 2 \end{bmatrix} \begin{matrix} =-x_5 \\ =-x_6 \\ \to \max \end{matrix} \end{array} \qquad \text{所有 } x_i \geqslant 0$$

解 此表不是标准型．为了将其化为标准型，将每行乘-1. 因为这样改变了目标函数，令 $f=-x_2+2x_3+2$. 新的目标函数是$-f$：

$$\begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & 1 \end{matrix} \\ \begin{bmatrix} -1 & -1 & 0 & -1 & 0 \\ 1 & 1 & 2 & 1 & -1 \\ 0 & 1 & -2 & 0 & -2 \end{bmatrix} \begin{matrix} =x_5 \\ =x_6 \\ =-f\to \min \end{matrix} \end{array} \qquad \text{所有 } x_i \geqslant 0$$

此表是标准的．因为它是不可行的，不能用第 2 阶段求解，只有等到学了第 1 阶段法才能求解吗？不．我们可以仔细考察这个问题，设法用普通方法求解．表中第 1 行“几乎是坏的”．它可写成$-x_1-x_2-x_4=x_5$. 因为所有的 $x_i\geqslant0$，我们可得 $x_1=x_2=x_4=0=x_5$. 于是只需确定 x_3 和 x_6 就可以了，为此我们有下面规模较小的标准表：

106

$$\begin{array}{c} \begin{matrix} x_3 & 1 \end{matrix} \\ \begin{bmatrix} 2 & -1 \\ -2 & -2 \end{bmatrix} \begin{matrix} =x_6 \\ =-f\to\min \end{matrix} \end{array}, \quad x_1=x_2=x_4=x_5=0;x_3,x_6\geqslant 0$$

此表仍是不可行的，但有一个坏列．因为不能用第 2 阶段的方法求解，所以我们不能得到问题无界的结论．但是，可以将第 2 阶段中有坏列时证明问题无界的讨论运用到这里．即 $x_3\geqslant1/2$ 时解是可行的，我们令 x_3 趋于∞，此时目标函数值趋于$-\infty$. 于是问题无界，$\min(-f)=-\infty$. 对于原问题，它也是无界的：$\max=\infty$. 另一种方法是用图解法求关于 x_3 和 x_6 的最优化问题． ■

从初始表开始，由单纯形法所能得到的表的数目是有限的：只有有限种方法可将一些变量放在顶端，其余的变量放在右端，变量的位置固定后，表也随之确定．注意所有写成这种表形式的线性方程组都是等价的(对初始表而言，这个方程组是 $Ax^{\mathrm{T}}=u$)，因此它们有同样的解．

事实上，如果把 m 个变量放在右端(基变量)，n 个变量放在顶端(非基变量)，那么就有$(m+n)!$ 种方法在 $m+n$ 个位置上安排 $m+n$ 个变量．当这些变量的位置确定后，便可确定$(m+1)\times(n+1)$的矩阵．这些表中有些是可行的，有些则不可行，因为我们的线性方程组对某些包含 m 个变量的集合而言是不可解的．如果我们改变基变量或非基变量的内部顺序，不会改变基解．因此基解的数目至多是$\dfrac{(m+n)!}{m!n!}$.

由单纯形法的第 2 阶段，要么经有限步得到最优表，要么得到一个前面已经出现的表．对于后一种情况，我们将进入一个循环，得到具有相同元素 d 的相同的表集合．注意，最后一行的最后一个元素 d 即当前的目标函数值在整个循环中保持不变，因为它不能增大．因为 d 不变，故在整个循环中旋转行的最后一个元素是 0.

107

若旋转行的最后一个元素是 0，则称此旋转步是*退化的*．等价地，一个退化的旋转步不能改变基解(虽然它确实能改变基变量的集合).

避免循环的一个好方法是，在第 2 阶段中，如果存在多个可供选择的行，我们随机地选取其中一个 i'. 有多种方法可以确定步骤 3 中旋转元的选取，故有多种单纯形算法．不能说哪种方法对所有的数据而言都是最好的，但有些方法可以避免出现循环．

还有一种选取旋转列的好方法是，在步骤 3 中取最小的 c_j(这称为最小系数法则). 但是，当存在好几个最小元素时，这种法则需要进一步精确化，以便成为一个精确算法．

我们也可以通过计算目标函数值的改变量来确定旋转元的选取，在所有保持可行性条件的旋转元中取能给出最好改变量(最大下降规则)的元素作为旋转元．然而这种法则与前面的法则有时可能是矛盾的．

实际问题中循环很少发生，因此通常并不需要设法防止循环．对单纯形法而言，还有另一些更重要的问题，我们很快将会讨论．但是，从理论上证明如何避免循环是很重要的．也就是说，设计一个一直能进行下去的单纯形法(这里“能进行下去”的意思是指在有限步内终止)是很重要的．下面我们讨论这个问题．

防止循环的第一种方法是*扰动法*(perturbation)，也就是将最后一列的元素做一些小改变使其不为零．更确切地说，用扰动列 $b(\varepsilon)=[b_1+\varepsilon, \cdots, b_n+\varepsilon^n]^T$ 代替列 $b=[b_1, \cdots, b_n]^T$. 正式地说，我们运用的是 ε 的多项式；非正式地说，$b(\varepsilon)$是 b 的一个小扰动，这里 ε 是一个小正数．我们将定义如何比较这些多项式，然后用单纯形法求解扰动矩阵．b 部分的所有元素保持非零，使得循环不可能发生．参见附录 A6.

虽然这种扰动法需要额外的计算量且实际应用价值不大，但是它可以说明存在避免循环的办法．

1976 年 Robert G. Bland 提出了一种反循环法则，使步骤 3 中的所有选择唯一．这里我们解释一下 Bland 法则．首先，旋转前，将变量名或者标识按行(或列)列出排序．例如，若有 3 个变量，则可以命名并排序为 x, y, z，或标识它们并排序为 t_1, t_2, t_3. 其次，只要单纯形法中需要进行变量选择，就按照排好的变量序列选取排在最前面的变量．若变量是按照下标排序的，该法则就称为最小下标法则．

108

例如，假设在最后一行有两个负的元素对应着变量 x 和 y，那么旋转列将是

x 对应的列，因为按照排序，x 出现在 y 的前面．类似地，假设有两个最大的比值对应着 x 和 y，那么 x 对应的行将是旋转行．

下面我们证明 Bland 法则是可行的．

定理 10.11 *单纯形法的第 2 阶段按照 Bland 法则选取旋转元不会产生循环．*

证明 设变量为 x_1，…，x_{n+m}，使用最小下标法则，我们在 $(m+1)\times(n+1)$ 的标准可行的行表上运算．

假若产生一个循环．**回转**(swing)变量是指在循环中位置发生变化的变量．当循环出现时，回转变量在一个表中标识旋转列而在另一个表中标识旋转行．

设 x_t 是一个下标最大的回转变量．考虑表

$$\begin{array}{cc} y & 1 \end{array}\qquad$$
$$\begin{bmatrix} A & b \\ c & d \end{bmatrix} \begin{array}{l} = z \\ = f \to \min \end{array} \tag{10.12}$$

这里 x_t 标识旋转列（即 x_t 在顶端，但它在下一个表的右端），则旋转列的最后一个元素 c_t 是负的，顶端变量 $x_j(j<t)$ 对应的 $c_j\geqslant 0$，否则根据 Bland 法则应选取另一个旋转列．注意回转行中的最后一个元素一定为零．

现在考虑表

$$\begin{array}{cc} y' & 1 \end{array}\qquad$$
$$\begin{bmatrix} A' & b' \\ c' & d' \end{bmatrix} \begin{array}{l} = z' \\ = f \to \min \end{array} \tag{10.13}$$

这里 x_t 标识旋转行．也就是说，x_t 在右端（z' 中的变量），但将被移到顶端．注意 $d=d'$，因为在循环中目标函数值不增加．

109

令

$$\begin{array}{c} x_s \\ \begin{bmatrix} a \\ c_s \end{bmatrix} \end{array}$$

是(10.13)中的旋转列．故有 $c_s<0$，$a_t<0$，且当 $x_i(i\neq t)$ 是右端的回转变量时，x_i 对应行中 a 的元素 $a_i\geqslant 0$.

下面我们确定方程组 $Ay^{\mathrm{T}}=z$ 或其等价的方程组 $A'y'^{\mathrm{T}}=b'$ 的一个解 h，它是不可行的，也不要求它是基解．令 $x_s=1$，y' 中的其他变量为 0，则有 $z'=a+b'$，$f(h)=d+c_s<d$. 特别是对此解而言，对顶端的任意变量 $x_j(j\neq s)$ 有 $x_j=h_j=0$；对(10.13)右端的任意回转变量 x_i 有 $x_i=h_i=a_i\geqslant 0$.

另一方面，可将此解加到表(10.12)中，得到 $f(h)=d+\sum c_jh_j$，其中和式取遍 y 中所有回转变量[回忆一下，表(10.12)和(10.13)顶端的非回转变量的值都是 0]. 在这个和式中，$c_th_t=c_ta_t>0$（因 c_t，$a_t<0$）且其余项都 $\geqslant 0$（因对 $j\neq t$ 有 c_j，$h_j>0$）. 于是有 $f(h)>d$，与前面的结果矛盾． ■

练习

1～3. 求解下面由标准行表给出的线性规划：

1.
$$
\begin{array}{ccccc|l}
a & b & c & d & 1 & \\
0.3 & 0.35 & 0.5 & 0.4 & 0.4 & =y_1 \\
-0.3 & -0.35 & -0.5 & 0.4 & 0.4 & =y_2 \\
0.6 & 0.35 & 0.5 & 0.45 & 0 & =y_3 \\
-0.6 & -0.35 & -0.5 & 5 & 0.5 & =y_4 \\
0.1 & 0.3 & 0 & 0.15 & 1 & =y_5 \\
-0.1 & -0.3 & 0 & -0.15 & 0.1 & =y_6 \\
11 & 12 & 16 & 14 & 0 & \to \min
\end{array}
$$

2.
$$
\begin{array}{ccccc|l}
y_7 & y_8 & y_9 & y_{10} & 1 & \\
0.3 & 0.35 & 0.5 & 0.4 & -0.4 & =y_1 \\
-0.3 & -0.35 & -0.5 & -0.4 & -0.4 & =y_2 \\
0.6 & 0.35 & 0.5 & 0.45 & -0.5 & =y_3 \\
-0.6 & -0.35 & -0.5 & -0.45 & 0.5 & =y_4 \\
0.1 & 0.3 & 0 & 0.15 & -0.1 & =y_5 \\
-0.1 & -0.3 & 0 & -0.15 & 0.1 & =y_6 \\
11 & 12 & 16 & -14 & 0 & =C \to \min
\end{array}
$$

110

3.
$$
\begin{array}{ccccc|l}
z_1 & z_2 & z_3 & z_4 & 1 & \\
0 & 0.35 & 0.5 & 0.4 & 0.4 & =y_1 \\
3 & -0.35 & -0.5 & -0.4 & 0.4 & =y_2 \\
0.6 & 0.35 & 0.5 & 0.45 & 0 & =y_3 \\
0.6 & -0.35 & -0.5 & -0.45 & 0.5 & =y_4 \\
0.1 & 0.3 & 0 & 0.15 & -0.1 & =y_5 \\
0 & -0.3 & 0 & -0.15 & 0.1 & =y_6 \\
-11 & 12 & 16 & 14 & 0 & \to \min
\end{array}
$$

4～6. 对以下陈述回答是或否．

4. 标准表是最优的，若它的基解是最优解．

5. 只有一行的标准表是可行的．

6. 若一个标准表仅有一行且无坏列，则它是最优的．

7. 求解例 2.3 给出的线性规划．

8. 解线性规划 $x+z\to\min$，$y-u=1$，$x\geqslant 0$，$y\geqslant 0$.

9. 求解

$$1.2a+1.4b+1.7c+1.9d\to\min, a+b+c+d=1,$$
$$0.1a+0.2b+0.3c+0.4d=0.25, [a,b,c,d]\geqslant 0.$$

10. 求解

$$\begin{array}{ccccc} x_1 & -x_2 & x_3 & x_4 & -1 \\ \end{array}$$
$$\begin{bmatrix} 1 & 1 & 0 & 1 & -3 \\ -1 & -1 & -2 & -1 & 1 \\ 0 & -1 & 2 & 0 & 2 \end{bmatrix} \begin{array}{l} =x_5 \\ =-x_6 \\ \to\min \end{array} \qquad \text{所有 } x_i\geqslant 0$$

11. 求解

$$\begin{array}{ccccc} x_1 & -x_2 & x_3 & x_4 & -1 \\ \end{array}$$
$$\begin{bmatrix} 1 & -1 & 0 & 1 & -3 \\ -1 & 1 & -2 & -1 & 1 \\ 0 & -1 & 2 & 0 & 2 \end{bmatrix} \begin{array}{l} =x_5 \\ =-x_6 \\ \to\max \end{array} \qquad \text{所有 } x_i\geqslant 0$$

12. 求解

$$\begin{array}{ccccc} x_1 & -x_2 & x_3 & x_4 & 1 \\ \end{array}$$
$$\begin{bmatrix} 1 & -1 & 0 & 1 & -3 \\ -1 & 1 & -2 & -1 & 1 \\ 0 & -1 & 2 & 0 & 2 \end{bmatrix} \begin{array}{l} =x_5 \\ =-x_6 \\ \to\max \end{array} \qquad \text{所有 } x_i\geqslant 0$$

13. 描述第 2 阶段法如何求解仅有一行的标准表．提示：首先考虑几个例子．

14. 描述第 2 阶段法如何求解仅有两列的可行表(即顶端只有一个变量)．提示：首先考虑几个例子并用图解法．

111

11. 单纯形法的第 1 阶段

一个好的单纯形法对任意一个有最优解的线性规划都能找到一个最优表．对无最优解的问题，任何方法都不可能找到最优表．第 1 阶段的目的是找一个可行解，因为有些线性规划是不可行的，我们需要考虑单纯形法可以判别不可行的线性规划．回忆一下，单纯形法是在标准表上实施的．

定义 11.1 标准表中的坏行(bad row)是指最后一个元素严格为负，其他元素非正且不是最后一行的行．

换言之，它就是矩阵$[A, b]$中满足 $A_i\leqslant 0$，$b_i<0$ 的行$[A_i, b_i]$(见式(10.1))．与坏行对应的约束是 $A_i^{T}x+b_i=u_i$．式子左边是严格负的而右边是非负的，所以是矛盾的．于是由具有坏行的标准表表示的线性规划无可行解，这是因为坏行和非负条件不相容．

例如，式(10.4)中第 2 个表中 x 对应的行是一个坏行，而第 1 个表没有坏行．“问题不可行”简记为 $\min=\infty$．

从定义中容易得到线性规划和它的标准表之间如下的逻辑关系：

$$\text{有坏行的表} \Rightarrow \text{问题不可行} \Rightarrow \text{无可行表}. \qquad (11.2)$$

注意，存在不可行表不能说明 min=∞，而且 min=∞不能说明在每个表中都有坏列．

单纯形法的第 1 阶段将生成一个可行表或者一个有坏行的表，因此可以证明线性规划或者不可行(第二种情况)或者有一个可行解(即第一种情况的可行表的基解)．由此，我们有式(11.2)中逆向的蕴涵关系：

$$\text{有坏行的表} \Leftrightarrow \text{问题不可行} \Leftrightarrow \text{无可行表}.$$

112

综上所述，我们得到下面的定理：

定理 11.3 单纯形法可以得到并且只能得到以下三种结果之一：

1. 最优表；此时从最优表中可得最优解和最优值．
2. 有坏列的可行表；此时问题无界，因此无最优解．
3. 有坏行的表；此时问题不可行(即问题无可行解)．

由此可知单纯形法有三种不同类型的终止表．加上第 1 阶段的终止表——行可行表，是四种不同类型．辨别这些终止表很重要，可以知道算法何时停止并写出答案．下面的草图可帮助我们记忆终止表：

行可行表： $\begin{array}{c} \begin{matrix} \oplus & 1 \end{matrix} \\ \begin{bmatrix} * & \oplus \\ * & * \end{bmatrix} \end{array} \begin{array}{l} =\oplus \\ \to \min, \end{array}$ 终止第 1 阶段，转入第 2 阶段．

坏行： $\begin{array}{c} \begin{matrix} \oplus & 1 \end{matrix} \\ \begin{bmatrix} \ominus & - \end{bmatrix} \end{array} =\oplus,$ 终止表，min=∞.

最优表： $\begin{array}{c} \begin{matrix} \oplus & 1 \end{matrix} \\ \begin{bmatrix} * & \oplus \\ \oplus & * \end{bmatrix} \end{array} \begin{array}{l} =\oplus \\ \to \min, \end{array}$ 终止表，写出答案．

坏列： $\begin{array}{c} \oplus \\ \begin{bmatrix} \oplus \\ - \end{bmatrix} \end{array} \begin{array}{l} =\oplus \\ \to \min, \end{array}$ 若在第 2 阶段，则终止．

有坏列的可行表： $\begin{array}{c} \begin{matrix} \oplus & \oplus & \oplus & 1 \end{matrix} \\ \begin{bmatrix} * & \oplus & * & \oplus \\ * & - & * & * \end{bmatrix} \end{array} \begin{array}{l} =\oplus \\ \to \min, \end{array}$ 终止，min=−∞.

这里：⊕代表正元和零元，即非负元(≥0)；⊖代表负元和零元，即非正元(≤0)；− 代表负元；+ 代表正元；* 代表任意数．

113

在第 10 节我们给出了从可行标准表中选择旋转元的方法．然而，当给出的标准表不可行时怎么办呢？我们将给出一种生成可行标准表的方法或者证明这是不可能的．考虑标准表

$$\begin{array}{c} \begin{matrix} x & 1 \end{matrix} \\ \begin{bmatrix} A & b \\ c & d \end{bmatrix} \end{array} \begin{array}{l} = u \\ = z \to \min, x \geqslant 0, u \geqslant 0 \end{array}$$

像第 2 阶段一样，我们将包含一个旋转步的过程分成 4 小步．

单纯形法的第 1 阶段

1. 若 $b \geqslant 0$，则我们的表已经可行，第 1 阶段完成，进入第 2 阶段．

2. 检验表中是否有坏行．也就是，对每个 $b_i < 0$，检验它所在行中的元素 a_{ij} 的符号．若有一个这样的 b_i，它所在行中的元素 a_{ij} 都非正，则我们不能满足此约束．因此，问题无可行解．我们可写出如下答案：

$$\min = \infty$$

3. 选取最靠近表的顶端且满足 $b_{i'} < 0$ 的行 i'，这意味着对于 $b_{i'}$ 上面的行 i 有 $b_i \geqslant 0$. 在该行中选取任意正元 $a_{i'j'} > 0$. 取 $i'' \leqslant i'$ 使得 $b_{i''}/a_{i''j'} = \max(b_{i'}/a_{i'j'},\ b_i/a_{ij'})$，这里对所有的满足 $a_{ij'} < 0$ 的指标 $i \leqslant i'$ 取最大．

4. 旋转，转向步骤 1.

注意位于 $b_{i'}$ 上面的 b 元素保持非负，而 $b_{i'}$ 的值要么增大要么保持不变．事实上，$b_i (i \leqslant i')$ 被替换为 $\overline{b_i} = b_i - a_{ij'} \cdot b_{i''}/a_{i''j'}$. 若 $a_{ij'} \geqslant 0$（例如，$i = i'$ 时），则 $\bar{b}_i \geqslant b_i$. 若 $a_{ij'} < 0$，则 $\overline{b_i} \geqslant 0$.

上述方法或者生成一个有坏行的表，或者生成一个可行表，或者进入一个循环．仅当 b 所在列有零元且那些零元位于旋转行（退化情况）时，才可能出现循环．为防止出现循环，可以采用 Bland 法则（见第 10 节）或者使用扰动法选择旋转元. Bland 法则是有效的，因为如果我们进入一个循环，那么存在一个线性规划，我们在第 2 阶段也遇到循环．这个规划是通过改变标准表第一行中的最后一个负元素的符号得到的，将其当作目标函数处理而忽略其他行．

114

实际计算表明，一个好的单纯形法的旋转步数与 $\min(m,\ n)$ 成比例，其中 $m \times n$ 是表的大小．但是，对许多用来确定旋转元（包括 Bland 法则）的常用方法而言，可以构造一个线性规划问题，需要比此多得多的旋转步．幸运的是，“病态的”问题在实际中一般不会出现．

对任意自然数 m，n 及任意 $m+1$ 行 $n+1$ 列的标准可行表（即有 $m+n$ 个决策变量），是否存在某个数 t，对可行表经过少于 $(m+n)^t$ 次的旋转后（不一定用单纯形法）生成终止表，这仍是一个尚未解决的问题．去掉表可行的限制，可取 $t = 1$，因为我们可以经过 $\leqslant \min(m,\ n)$ 次旋转步得到终止表．

如果用计算机，除了需要数据的储存空间外，还需要当前表的储存空间．由于舍入误差的积累，偶尔可能需要重新从初始表开始计算当前表（这是要保存初始表的原因）. 这需要 $\leqslant \min(m,\ n)$ 次旋转．否则，积累误差可能使你远离最优可行解．一般地，时间就是金钱，所以偶尔检查一下问题的进展（即目标函数值的改进），确定是否值得继续花费时间和精力是一个好的想法．

大规模的线性规划通常有稀疏表，即系数矩阵的大多数元素是 0. 仅在内存中储存非零元素和使用乘法单纯形法能节省内存，乘法单纯形法是指用初始表的几个基本操作的乘积表示当前矩阵．

115

问题 11.4 求解

$$\begin{array}{cccc} x & y & z & 1 \end{array}$$
$$\begin{bmatrix} 1 & -1 & 1 & 2 \\ 0 & -1 & 0 & -1 \\ 1 & 0 & 2 & 0 \end{bmatrix} \begin{matrix} = u \\ = v \\ = w \to \min \end{matrix} \qquad x,y,z,u,v \geqslant 0$$

解 此表是标准的但不是可行的且 v 行是坏行，故问题不可行($\min=\infty$).

■

问题 11.5 求解

$$\begin{array}{cccc} x & y & z & 1 \end{array}$$
$$\begin{bmatrix} 1 & -1 & 1 & 2 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 2 & 0 \end{bmatrix} \begin{matrix} = u \\ = v \\ = w \to \min \end{matrix} \qquad x,y,z,u,v \geqslant 0$$

解 此表是标准的但不是可行的且无坏行．我们在第 2 行寻找一个正元，仅有一个在 y 列，所以将其作为旋转列．因 $2/(-1)<(-1)/1<0$；故 v 行是旋转行．进行旋转得到

$$\begin{array}{cccc} x & y & z & 1 \end{array}$$
$$\begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 \end{bmatrix} \begin{matrix} = u \\ = y \\ = w \to \min \end{matrix} \qquad x,y,z,u,v \geqslant 0$$

正如所料，表格保持标准型且最后一列的第一个元素非负．而且，此表是最优的．答案：$x=v=z=0$，$y=u=1$，$\min=0$. ■

问题 11.6 用单纯形法解问题(10.4).

解 得到标准型表

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} -1 & -1 & 0 & -1 & 0 \\ 1 & 1 & 2 & 1 & -1 \\ 1 & 0 & -2 & 0 & -2 \end{bmatrix} \begin{matrix} = x_5 \\ = x_6 \\ =-f \to \min \end{matrix} \qquad 所有\ x_i \geqslant 0$$

和以前一样，使用第 2 阶段法．我们看一下是否得到同样的答案．

有四种旋转列的选择结果和第 1 阶段法一致．它们中只有 x_3 列给出非退化的旋转步．用 2 做旋转元，旋转得到

$$\begin{array}{ccccc} x_1 & x_2 & x_6 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} -1 & -1 & 0 & -1 & 0 \\ -1/2 & -1/2 & 1/2 & -1/2 & 1/2 \\ 1 & 2 & -1 & 1 & -3 \end{bmatrix} \begin{matrix} = x_5 \\ = x_3 \\ =-f \to \min \end{matrix} \qquad 所有\ x_i \geqslant 0$$

此表是可行的，可使用第 2 阶段法．结果有一个坏列，问题无界．为了得到更多的练习，尝试一下其他三个旋转元．会进入循环吗？很有可能！ ■

问题 11.7 解线性规划

$$\begin{array}{c} \begin{matrix} x & y & 1 \end{matrix} \\ \begin{bmatrix} -1 & 1 & 1 \\ 2 & -1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{matrix} \geqslant 0 \\ \leqslant 0 \\ = z \to \max \quad x, y \geqslant 0 \end{matrix} \end{array}$$

解 此表不是标准型．后两行乘－1 并引入两个松弛变量 u，v，将其化为标准型

$$\begin{array}{c} \begin{matrix} x & y & 1 \end{matrix} \\ \begin{bmatrix} -1 & 1 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{matrix} = u \\ = v \\ = -z \to \min \quad x, y \geqslant 0, u, v \geqslant 0 \end{matrix} \end{array}$$

此表是最优的．故基解就是最优解：$x=y=0$，$u=1$，$v=0$，$\min=-1$．原问题答案是：$x=y=0$，$\max=1$．所有的最优解为 $x=0$，$y\leqslant 1$． ■

问题 11.8 求解由下面的标准行表给出的线性规划：

$$\begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & 1 \end{matrix} \\ \begin{bmatrix} 1 & 0 & -1 & -1 \\ -1 & 1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & -1 & 1 & 0 \end{bmatrix} \begin{matrix} = x_4 \\ = x_5 \\ = x_6 \\ \to \min \end{matrix} \end{array}$$

解 此表不可行．由第 1 阶段法，需要交换 x_1 和 x_4，旋转后得到

$$\begin{array}{c} \begin{matrix} x_4 & x_2 & x_3 & 1 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & -1 & 2 & 1 \end{bmatrix} \begin{matrix} = x_1 \\ = x_5 \\ = x_6 \\ \to \min \end{matrix} \end{array}$$

117

交换 x_2 和 x_5 得

$$\begin{array}{c} \begin{matrix} x_4 & x_5 & x_3 & 1 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ -1 & -1 & 0 & 0 \\ 0 & -1 & 2 & 0 \end{bmatrix} \begin{matrix} = x_1 \\ = x_2 \\ = x_6 \\ \to \min \end{matrix} \end{array}$$

现在表是可行的，可用第 2 阶段法．交换 x_5 和 x_6，做一个退化旋转步后得到一个最优表

$$\begin{array}{c} \begin{matrix} x_4 & x_6 & x_3 & 1 \end{matrix} \\ \begin{bmatrix} * & * & * & 1 \\ * & * & * & 1 \\ * & * & * & 0 \\ 1 & 1 & 2 & 0 \end{bmatrix} \begin{matrix} = x_1 \\ = x_2 \\ = x_5 \\ \to \min \end{matrix} \end{array}$$

因此 min=0，在 $x_4=x_6=x_3=0$，$x_4=1$，$x_2=1$，$x_5=0$ 达到． ■

注 在一些教材和一些计算机上实现的算法中，使用单纯形法从第 1 阶段过渡为第 2 阶段有一些技巧，而不是将单纯形法的两个阶段进行等同的处理．下面我们给出一个这样的技巧．考虑一个标准行表

$$\begin{array}{c} \begin{matrix} x & 1 \end{matrix} \\ \begin{bmatrix} A & b \\ c & d \end{bmatrix} \begin{matrix} = u \\ = z \to \min, x \geqslant 0, u \geqslant 0 \end{matrix} \end{array} \tag{11.9}$$

它不可行(即 b 的最小元 $-\mu$ 是负的)．引入新的变量 t，考虑标准行表

$$\begin{array}{c} \begin{matrix} x & t & 1 \end{matrix} \\ \begin{bmatrix} A & I & b \\ 0 & 1 & 0 \end{bmatrix} \begin{matrix} = u \\ \to \min \end{matrix} \quad \begin{matrix} x \geqslant 0, t \geqslant 0 \\ u \geqslant 0 \end{matrix} \end{array} \tag{11.10}$$

其中 I 是全部由 1 组成的列．以在 t 列且旋转行的最后元 $-\mu$ 为旋转元做一步旋转，得到可行表，因此一步旋转就可完成第 1 阶段．如果(11.10)的最优值是 0，则由(11.10)的最优表可容易地得到(11.9)的可行表．如果(11.10)的最优值不是 0，则问题(11.9)不可行． [118]

这种技巧的一种改进是"大 M 法"．用$[c, M, d]$代替(11.10)的最后一行$[0, 1, 0]$，这里 M 是一个很大的数．通过相同的旋转步，M 表成为可行的．若 M 规划有一个最优解且 $t=0$，则去掉 t 就可得到(11.9)的一个最优解．

还有其他避免或缩短第 1 阶段的技巧，其中有些是对 u 中的每个变量都引入一个新变量．

注 将第 2 阶段转化为第 1 阶段也总是可能的．这种做法的简单思想是对不同的 M 用线性约束 $f \leqslant M$ 代替目标函数 z，计算量可能很大．另一种方法是使用对偶，将在下一章介绍．

练习

1. 求解

$$\begin{array}{c} \begin{matrix} x & y & z & 1 \end{matrix} \\ \begin{bmatrix} 1 & -1 & 1 & 2 \\ 0 & -1 & 0 & -1 \\ 1 & 0 & 2 & 0 \end{bmatrix} \begin{matrix} = u \\ = v \\ = w \to \min \end{matrix} \quad \begin{matrix} x, y, z \geqslant 0 \\ u, v \geqslant 0 \end{matrix} \end{array}$$

2. 求解

$$\begin{array}{c} \begin{matrix} x & y & z & 1 \end{matrix} \\ \begin{bmatrix} 1 & -1 & 1 & 2 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 2 & 0 \end{bmatrix} \begin{matrix} = u \\ = v \\ = w \to \min \end{matrix} \end{array}$$

且 $x\geqslant 0$，$y\geqslant 0$，$z\geqslant 0$，$u\geqslant 0$，$v\geqslant 0$.

3. 求解

$$\begin{array}{c} \begin{matrix} x & y & z & 1 \end{matrix} \\ \begin{bmatrix} 1 & -1 & 1 & 2 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & -2 & 0 \end{bmatrix} \begin{matrix} = u \\ = v \\ = w \to \min \end{matrix} \end{array}$$

且 $x\geqslant 0$，$y\geqslant 0$，$z\geqslant 0$，$u\geqslant 0$，$v\geqslant 0$.

4～8. 求解下列线性规划，其中所有决策变量都要求非负．

4. $$\begin{array}{c} \begin{matrix} a & -b & c & -2 \end{matrix} \\ \begin{bmatrix} 1 & 0 & -1 & 2 \\ 2 & -1 & 0 & -3 \\ 0 & 2 & -1 & 0 \end{bmatrix} \begin{matrix} = d \\ = e \\ \to \min \end{matrix} \end{array}$$

5. $$\begin{array}{c} \begin{matrix} a & b & c & 1 \end{matrix} \\ \begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & 0 & 1 & 3 \\ -1 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} = d \\ = e \\ \to \min \end{matrix} \end{array}$$

119

6. $$\begin{array}{c} \begin{matrix} a & b & c & -1 \end{matrix} \\ \begin{bmatrix} 1 & 2 & 3 & -3 \\ 2 & 1 & 0 & -1 \\ -3 & 0 & 1 & 0 \end{bmatrix} \begin{matrix} = d \\ = e \\ \to \min \end{matrix} \end{array}$$

7. $$\begin{array}{c} \begin{matrix} a & b & c & -1 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} \begin{matrix} = c \\ = d \\ = f \to \min \end{matrix} \end{array}$$

8. $$\begin{array}{c} \begin{matrix} a & b & c & -1 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} \begin{matrix} = e \\ = d \\ = f \to \min \end{matrix} \end{array}$$

9～10. 下面的说法正确吗?

9. 不存在有一个坏行的标准可行表.

10. 不存在既有一个坏行又有一个坏列的标准表.

11～13. 求解下面由标准行表给出的线性规划.

11.
$$\begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & 1 \end{array}$$
$$\left[\begin{array}{rrrrr} 1 & 0 & -2 & -3 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ 2 & -1 & 0 & 1 & 3 \\ 1 & -1 & 1 & 0 & 2 \end{array}\right]\begin{array}{l} =x_5 \\ =x_6 \\ =x_7 \\ \to \min \end{array}$$

12.
$$\begin{array}{cccc} x_1 & x_2 & x_3 & 1 \end{array}$$
$$\left[\begin{array}{rrrr} 1 & 0 & -1 & -1 \\ -1 & 3 & 1 & 0 \\ 3 & -1 & 2 & 1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 0 \end{array}\right]\begin{array}{l} =x_4 \\ =x_5 \\ =x_6 \\ =x_7 \\ \to \min \end{array}$$

13.
$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & 1 \end{array}$$
$$\left[\begin{array}{rrrrrr} 1 & 0 & -1 & -1 & 0 & 1 \\ -1 & 3 & 1 & 0 & -2 & -1 \\ 3 & -1 & 2 & 1 & 2 & 1 \\ 1 & -1 & 1 & 0 & -1 & 1 \\ 0 & -1 & -1 & 0 & -1 & 2 \end{array}\right]\begin{array}{l} =x_6 \\ =x_7 \\ =x_8 \\ =x_9 \\ \to \min \end{array}$$

[120]

12. 几何解释

本节我们解释单纯形法第 2 阶段的几何意义，目的是提高我们的几何想象力，但代价是需要学习一些几何语言：n 元数组是 R^n 中的点，线是两个 n 元数组的线性组合，可行域是 R^n 中具有有趣性质的子集等等．单纯形法的第 2 阶段是在可行域中移动的一种方法．我们首先看一下作为 R^n 子集的可行域的性质．

由第 3 节(第 1 章)可知两个或更少变量的线性规划的可行域是一个具有非常特殊的性质的集合．事实上，当我们设法将一维的射线、区间推广到高维空间时，就可导出下面定义的凸集的概念．

定义 12.1 称集合 S 是凸的(convex)，如果对任意 x 和 y 属于 S，$0\leqslant a\leqslant 1$，有 $(1-a)x+ay$ 属于 S.

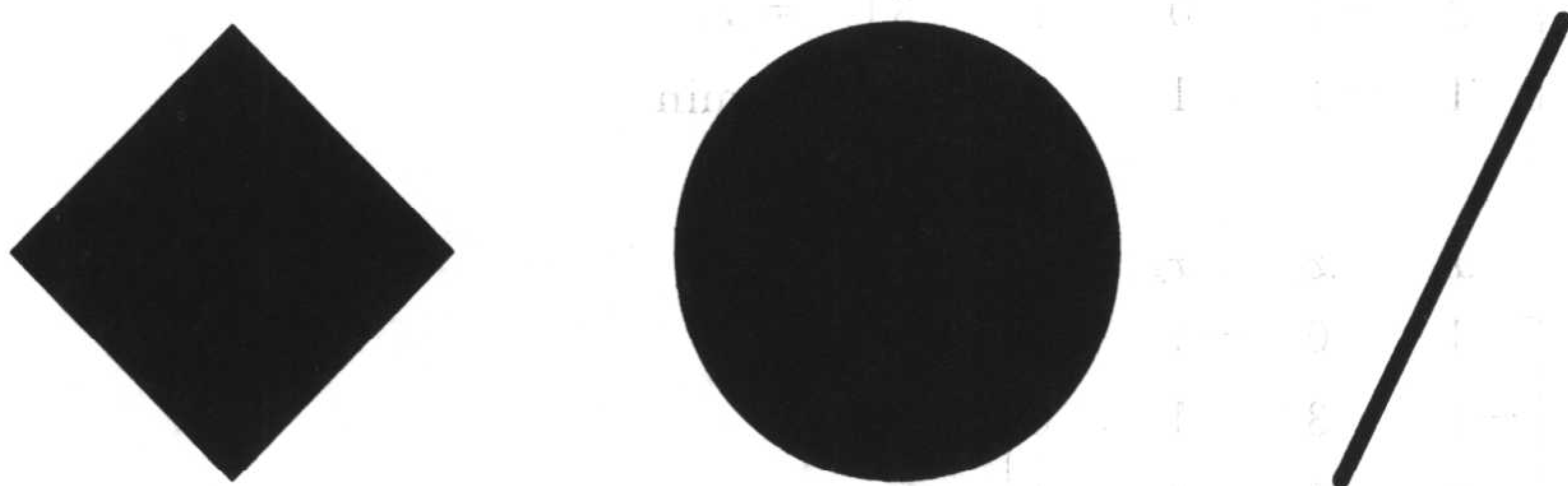
要使凸集的定义合理，$(1-a)x+ay$ 必须有定义．例如，当 $n\geqslant 1$ 是一个固定的整数，而 S 是行 $[a_1, a_2, \cdots, a_n]$ 或列 $[a_1, a_2, \cdots, a_n]^T$ 的集合时没有问

题．显然，我们可将这些行(列)与数相乘，然后相加得到新的行(列)．

定义 12.2 对 $0 \leqslant a \leqslant 1$，点 $(1-a)x+ay$ 称为 x 和 y 的凸组合(convex combination)或混合(mixture)．

两个不同点 x 和 y 的凸线性组合构成的集合称为连接 x 和 y 的区间或线段，也称为 x 和 y 的凸包(convex hull)．点 x 和 y 称为线段的端点．有时这些名称在 $x=y$ 时也使用，此时线段变成了一个点．

图 12.3 给出了三个凸集的例子．



121

图 12.3 平面上的三个凸集

其中包括：菱形 $|x|+|y| \leqslant 1$；圆盘 $(x-3)^2+y^2 \leqslant 1$；区间 $(1-\alpha)[5,-1]+\alpha[6,1](0 \leqslant \alpha \leqslant 1)$，连接着点 $x=5$，$y=-1$ 和 $x=6$，$y=1$．

图 12.4 给出了三个非凸集的例子．

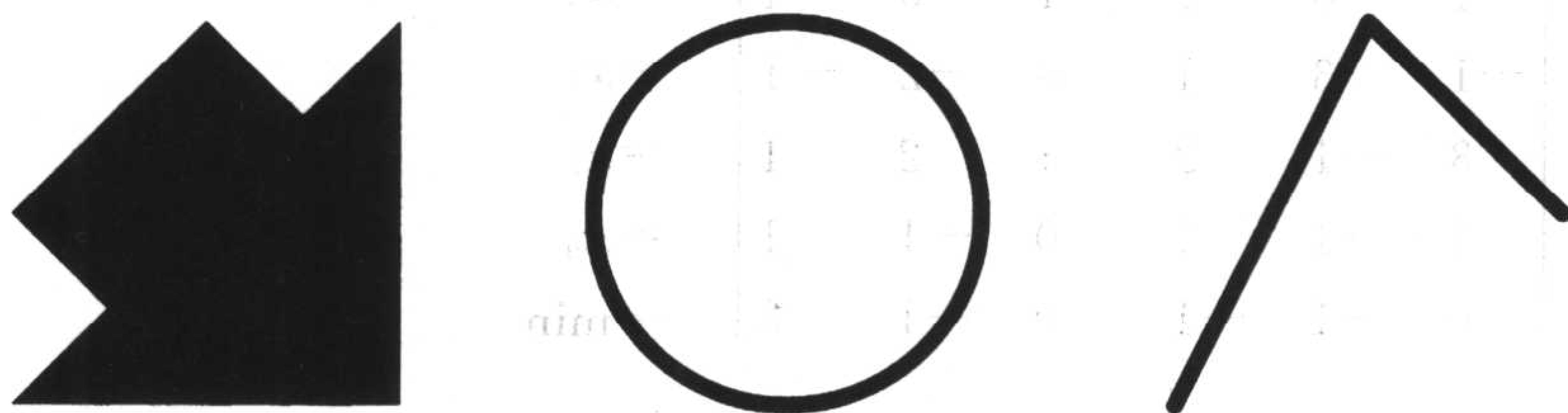


图 12.4 平面上的三个非凸集

其中包括：正方形和三角形的组合；圆周 $(x-3)^2+y^2=1$；两个区间的并集．

问题 证明圆盘 $x^2+y^2 \leqslant 1$ 是 (x, y) 平面上的凸集．

解 设 (x_1, y_1) 和 (x_2, y_2) 是圆盘中的两个元素，则有 $x_1^2+y_1^2 \leqslant 1$ 和 $x_2^2+y_2^2 \leqslant 1$．因为对 $0 \leqslant a \leqslant 1$，$(1-a)(x_1, y_1)=((1-a)x_1, (1-a)y_1)$ 且 $a(x_2, y_2)=(ax_2, ay_2)$，故点 $(1-a)(x_1, y_1)+a(x_2, y_2)$ 的坐标为 $x=(1-a)x_1+ax_2$ 和 $y=(1-a)y_1+ay_2$．将 x，y 的表达式代入，我们需要证明 $x^2+y^2 \leqslant 1$．为了证明这一点，只需要利用下面的结果：由 $(x_1-x_2)^2 \geqslant 0$，得 $2x_1x_2 \leqslant x_1^2+x_2^2$． ■

命题 12.5 凸集的交集还是凸集．

证明 设 S_i，$i \in I$ 是一组凸集，需证其交集 $S=\bigcap_{i \in I} S_i$ 是凸的．

任取 x 和 y 属于 S，则对每个 $i \in I$，有 x 和 y 属于 S_i 因此由凸集的定义，

$(1-a)x+ay$ 属于 S_i，$i\in I$. 于是有 $(1-a)x+ay\in S$，这证明了 S 的凸性. ■

由下面两个定理容易看到凸集在线性规划中的重要性. [122]

定理 12.6 线性规划的可行域 S 是凸集.

证明 设 S 是线性约束的有限集的交集(intersection). 只需证明由一个线性约束给出的区域是凸的.

设对变量 z_1，…，z_n 的线性约束如下：

$$a_1z_1+\cdots+a_nz_n\leqslant c$$

(=或≥的情况可以与≤类似处理). 在区间 $0\leqslant a\leqslant 1$ 中任取一个 a，令 $x=(x_1,\cdots,x_n)$ 和 $y=(y_1,\cdots,y_n)$ 满足此线性约束，即

$$a_1x_1+\cdots+a_nx_n\leqslant c, a_1y_1+\cdots+a_ny_n\leqslant c.$$

取它们关于 a 和 $(1-a)$ 的线性组合，得到

$$a_1(ax_1+(1-a)y_1)+\cdots+a_n(ax_n+(1-a)y_n)\leqslant c;$$

即 $(1-a)x+ay$ 满足此线性约束. ■

因为线性方程是线性约束，定理 12.6 表明线性方程组的解集是凸的. 事实上对线性方程组此时有更强的结论：经过两个不同的解 x，y 的直线 $\{(1-a)x+ay\}$ 都是解，不只是线段 $\{(1-a)x+ay: 0\leqslant a\leqslant 1\}$. 点 $(1-a)x+ay$ 称为 x，y 的仿射组合(affine combination).

作为凸性的例子，n 维单位单纯形 $x_1+\cdots+x_{n+1}=1$，所有 $x_i\geqslant 0$ 是凸的. 当 $n=1$ 时，这恰是单位区间 $0\leqslant x_1\leqslant 1$，二维单纯形是一个三角形.

定理 12.7 线性规划的最优解集是凸集.

证明 由定理 12.6 可以得证，因为最优解集可以看作可行域再加一个线性约束

$$\text{目标函数}=\text{最优值}$$

的集合. ■

除了可行域和最优解集之外，还有一个与线性规划有关的第三个凸集，即可行值的集合. 下面给出可行值的定义. [123]

定义 12.8 优化问题在其可行域上的目标函数值称为可行值(feasible value).

如果可行域是空集(即约束系统不相容)，则该优化问题无可行值.

命题 12.9 凸集在仿射变换下的像集仍是凸集.

证明 设 S 是 n 元列向量组成的空间中的凸集，$x\mapsto Cx+d$ 是一个仿射变换，其中 C 是一个 $m\times n$ 常数矩阵，d 是一个 m 元常数列向量. 若 $Cx+d$ 和 $Cy+d$ 是像集中的两点，则对属于区间 $0\leqslant a\leqslant 1$ 的任意 a，有 $a(Cx+d)+(1-a)(Cy+d)=C(ax+(1-a)y)+d$ 也属于像集. ■

这个证明的要点是仿射变换下直线变为直线，线段变为线段.

推论 12.10 线性规划的可行值的集合是凸的.

证明 因为线性规划的可行集是凸集且目标函数是仿射变换，故由命题 12.9 即可得证. ■

定义 12.11 S 的顶点(vertex，或称为极点(extreme point))是指 S 中的一个点，它不能表示成 S 中两个不同的点 y，z 的半和$(y+z)/2$.

换言之，极点就是完全位于 S 中的线段的端点. 另一种说法是点 x 是 S 的顶点当且仅当去掉 x 后 S 仍是凸的. 顶点的另一种定义参见练习第 14 题.

例 12.12 区间 $0\leqslant x\leqslant 1$ 恰有两个顶点，就是端点 $x=0$ 和 $x=1$. 事实上，若有区间 $0\leqslant x\leqslant 1$ 中的 y 和 z 使得 $0=(y+z)/2$，则显然有 $y=z=0$. 同样地，1 是顶点. 另一方面，若 x 属于区间 $0\leqslant x\leqslant 1$ 且 $x\neq 0$，则 x 可表示为 $x=(y+z)/2$ 且 $y=x-\varepsilon$，$z=x+\varepsilon$，其中 $\varepsilon=\min(x,\ 1-x)\neq 0$，因此 y 和 z 都属于区间 $0\leqslant x\leqslant 1$ 且不相同. ■

例 12.13 证明三角形 $x\geqslant 0$，$y\geqslant 0$，$x+y\leqslant 1$ 有三个顶点：$x=y=0$；$x=0$，$y=1$；$x=1$，$y=0$. ■

124 **例 12.14** 证明圆盘 $x^2+y^2\leqslant 1$ 的极点是圆 $x^2+y^2=1$ 上的点. ■

顶点在仿射映射下的像不一定是顶点. 例如，取图 12.3 中有 4 个顶点的菱形 S，投影到 x 轴：$[x,\ y]^{\mathrm{T}}\mapsto x=[1,\ 0][x,\ y]^{\mathrm{T}}$. 它的像是有两个顶点(端点)的区间，$S$ 的另两个顶点变为了区间的中点.

命题 12.15 设 S 是凸集，f 是一个仿射变换，将 S 中的不同点变为不同点，则对 S 的任意顶点 x，$f(x)$是 $f(S)$的顶点.

证明 设 x 是 S 的顶点且存在 $f(S)$中的 $f(y)$和 $f(z)$使得 $f(x)=(f(y)+f(z))/2$. 则 $x=(y+z)/2$，故有 $x=y=z$. 因此 $f(x)=f(y)=f(z)$. 于是 $f(x)$是顶点. ■

例如，对所有变量均大于等于 0 的线性规划 P，可将其表示为一个标准表，故 P 仿射等价于表上的问题 P′. 命题 12.15 表明 P 的可行域的顶点和 P′的可行域的顶点是一一对应的. 现在我们来看可行表的顶点.

定理 12.16 考虑由标准行表给出的线性规划，其可行域的点是顶点的充要条件是它是可行表的基解.

证明 考虑任意可行表

$$\begin{array}{c} \begin{matrix} x & 1 \end{matrix} \\ \begin{bmatrix} A & b \\ c & d \end{bmatrix} \begin{matrix} =u \\ =z\rightarrow\min \quad x\geqslant 0,u\geqslant 0 \end{matrix} \end{array} \tag{12.17}$$

其相应的基解是 $x=0$，$u=Ax^{\mathrm{T}}+b=b$. 可行性是指 $b\geqslant 0$. 假设$[0,\ b]$可用可行解$[y,\ Ay^{\mathrm{T}}+b]$，$[z,\ Az^{\mathrm{T}}+b]$表示为

$$[0,b]=([y,Ay^{\mathrm{T}}+b]+[z,Az^{\mathrm{T}}+b])/2.$$

因为 y，$z\geqslant 0$，我们得到 $y=z=0$；因此$[y,\ Ay^{\mathrm{T}}+b]=[z,\ Az^{\mathrm{T}}+b]=[0,\ b]$.

于是$[0,\ b]$是一个顶点．

反之，设$[x',\ Ax'^{\mathrm{T}}+b]$是可行域 S 的顶点(若 S 为空集，则不需要证明)．从任意一个可行表(12.17)开始，通过退化旋转将尽可能多的在顶点取值为零的变量放在表的顶端．

假设在表(12.17)中，我们不能将顶端非零的任何一个变量和右端为零的任何一个变量进行交换．若顶端变量的值都是 0，则$[x',\ Ax'^{\mathrm{T}}+b]=[0,\ b]$是基解，证毕．假设顶端有一个变量 x_0 取非零值 x_0'，我们将得到矛盾． 125

通过交换列，我们可以假设这是顶端的第一个变量；通过交换行，我们可以假设常数列的零(如果有的话)在这列的上端．现在我们的表为

$$
\begin{array}{c} \begin{matrix} x_0 & y & 1 \end{matrix} \\ \begin{bmatrix} A' & * & 0 \\ * & * & + \\ * & * & d \end{bmatrix} \end{array}
\begin{array}{l} \\ = v \\ = w \\ = z \to \min \quad x \geqslant 0, u \geqslant 0 \end{array}
\tag{12.18}
$$

其中+代表正元．因为我们不能通过一个旋转步将 x_0 与 v 中任意变量交换，则 v 在表中无 A'或者 $A'=0$. 现在我们将顶点稍微变化一点，用 $x_0-\varepsilon$ 和 $x_0+\varepsilon$ 代替 x_0 且保持 y'的值和顶端变量 y 的值相同．对充分小的 ε 我们得到两个不同的可行解

$$[[x_0'-\varepsilon, y'], A[x_0'+\varepsilon, y']^{\mathrm{T}}+b]$$

和

$$[[x_0'+\varepsilon, y'], A[x_0'+\varepsilon, y']^{\mathrm{T}}+b],$$

且顶点$[[x_0',\ y'],\ A[x_0',\ y']^{\mathrm{T}}+b]$是这两个点的和的一半，矛盾． ■

作为定理 12.16 的推论，我们得到最优表的基解是可行域的顶点．于是，如果线性规划有一个最优解，则可行域的一个顶点就是这个最优解．这个论述称为*角点原理*(corner principle)，这里角点是指顶点（极点).

定理 12.16 的另一个推论如下：

推论 *线性规划的可行域仅有有限个顶点．*

证明 设 x 是线性规划 P 的 n 个变量组成的列向量，m 是 P 中不等式的个数，这里没有符号限制．若要求规划 P 的 n 个变量$\geqslant 0$，则 P 仿射等价于一个由标准表给出的线性规划，故由定理 12.16 和命题 12.15 得到顶点数至多是

$$\frac{(m+n)!}{m!n!}=\binom{m+n}{m}=\binom{m+n}{n},$$ 126

其中 n 是 P 的变量个数，$m+n$ 是 P 中不等式的个数(注意将 P 化为一个标准表 P′时对每个无符号约束的不等式引入一个松弛变量). 一般地，我们将证明 P 的顶点数至多是$\binom{m+n}{m}$. 假设我们找到 P 的顶点的一个有限集合 V 并且顶点数 $N>\binom{m+2n}{m}$(我们并不知道 P 是否有有限个顶点)，则我们能替换全部变量，即

作变换 $x \mapsto y=x+s$（仿射变换反过来也是一个替换，即$x \mapsto x=x-s$），其中 s 是一个常数列向量，使得对选择的集合 V 中的所有顶点有 $y \geqslant 0$. 现在我们在列 y 的变量上加上至多 n 个符号限制得到一个有 n 个变量 $y \geqslant 0$ 的新问题 P″，它同样有 m 个另外的不等式．而且，V 中的顶点与 P″的顶点相对应．（P 的有些顶点现在可能被去掉了．）因有太多的顶点，我们得到矛盾．这里我们需要用到一个有关顶点的事实，参见这节后面的练习第 18 题． ■

由定理 12.16，在单纯形法的第 2 阶段，一个旋转步就是从可行域的一个顶点转到另一个顶点．每个顶点可由相应的基解或由线性方程组来描述：所有非基变量$=0$. 退化旋转步不能改变基解，但任何旋转步都改变基（右端变量的集合）和描述解的方程的集合．每次改变并不大：只有一个基变量被替换且只有一个方程被替换．

不难证明，被一个非退化旋转步关联的两个顶点在下面的意义下不同但相邻．

定义 12.19 *凸集 S 的两个不同的顶点 x，y 称为相邻的，如果删掉连接 x，y 的线段后 S 仍是凸的．*

事实上，假设 x 和 y 是可行表 T 和 T′的基解，T 和 T′由一个旋转步连接，且它们的凸组合 $z=ax+(1-a)y$ 可由可行解 u，v 表示成$(u+v)/2$. 我们需证明 u，v 都是 x 和 y 的凸组合．设旋转步交换了变量 x_1 和 x_2，则顶端的其他变量（非基变量）x'在两组顶点中都取为 0.

127

我们考虑表 T：

$$\begin{array}{ccc} x_1 & x' & 1 \end{array}$$
$$\begin{bmatrix} \alpha^* & * & \beta \\ * & * & * \\ * & * & * \end{bmatrix} \begin{array}{l} = x_2 \\ = x'' \\ \to \min \end{array}$$

这里为了方便，我们用 x_1 表示顶端第一个变量，x'表示顶端的其他变量，x_2 表示右端第一个变量，x''表示边上的其他决策变量．

因为变量 x'在 x，y 中取零，也在 z 取零，故在 u，v 也取零．因此 x，y，z，u，v 由 x_1 的值决定．x_1 的值对 x 和 y 分别是 0 和$-\beta/\alpha$. 这些值和它们之间的值（回忆在第 2 阶段旋转元 $\alpha<0$ 且 $\beta>0$）给出的是可行解，而其他值给出的是不可行解．于是 u，v 属于连接 x，y 的区间，它是 $0 \leqslant x_1 \leqslant -\beta/\alpha$ 在仿射变换下的像．

因此，一个退化旋转步不会使顶点改变而非退化旋转步使顶点改变为相邻的顶点．在这两种情况下，决定顶点的线性方程的集合换了一个方程：

非基变量$=0$.

我们仍然讨论由标准表(12.17)给出的线性规划

$$\text{minimize } z = cx^{\mathrm{T}} + d, \text{subject to } x \geqslant 0, Ax^{\mathrm{T}} = u \geqslant 0.$$

设顶端有 n 个变量，右端有 m 个变量．不等式 $x\geqslant 0$ 表示 R^n 中的 n 个半空间，$Ax^{\mathrm{T}}=u\geqslant 0$ 表示另外的 m 个半空间．故对 x 而言，R^n 空间的可行域是 $m+n$ 个半空间的交集．

有时可以更方便地认为可行域 S 是嵌入 R^{n+m} 的(x, u)空间的．此时，它由 $m+n$ 个不等式和 m 个等式 $Ax^{\mathrm{T}}=u$ 给出．

因为表是可行的，基解 $x=0$，$u=b$ 在 S 内，故 S 是非空集．它是非负维数的凸"多面体"或"多边形"(有界或无界．想知道更多关于有界集的知识，参见练习第 13 题)．

一般情况下(当 b 无零元时)，S 是 n 维的且有 n 个超平面(有效约束)通过基解，即令顶端变量为 0． [128]

在单纯形法的第 2 阶段，从一个顶点即初始表的一个基解开始．每个旋转步都变到相邻的具有相同或更好函数值的顶点(见定义 12.19)．在非退化情况，每个旋转步都严格改进目标函数值．在退化情况，通过一个有效约束代替另一个，实际上保持在相同的顶点．

得到一个坏列，意味着找到一个趋向无穷的"边"，使得沿着此边驱使目标函数值趋向 $-\infty$．仅当 S 无界时才会发生这种情况．

作为定义 12.8 的推论，我们可得最优表的基解是可行域的顶点．因此，若线性规划有一个最优解，则可行域的一个顶点就是这个最优解．这个论述称为*角点原理*，这里角点是指顶点(极点)．

例 12.20 下面考虑一个可画出图形的二维例子．设线性规划

$$x+8y\to\max, y-2x\leqslant 2, x+y\leqslant 5, 2x+y\leqslant 7, x\leqslant 3, x\geqslant 0, y\geqslant 0$$

用单纯形法求解．因约束是规范形，为得到标准行表，只需将目标函数乘 -1 并引入松弛变量 a，b，c，d：

$$\begin{array}{c} \begin{array}{ccc} x & y & 1 \end{array} \\ \begin{bmatrix} 2 & -1 & 2 \\ -1 & -1 & 5 \\ -2 & -1 & 7 \\ -1 & 0 & 3 \\ -1 & -8 & 0 \end{bmatrix} \begin{array}{l} =a \\ =b \\ =c \\ =d \\ =-f\to\min \end{array} \end{array}$$

此表可行，所以我们进入第 2 阶段．最后一行有两个负数，可任选一个进旋转列．选取 y 入基，最大比是 -2，故 u_1 出基．通过这个旋转步后得到表 [129]

$$\begin{array}{c} \begin{array}{ccc} x & a & 1 \end{array} \\ \begin{bmatrix} 2 & -1 & 2 \\ -3 & 1 & 3 \\ -4 & 1 & 5 \\ -1 & 0 & 3 \\ -17 & 8 & -16 \end{bmatrix} \begin{array}{l} =y \\ =b \\ =c \\ =d \\ =-f\to\min \end{array} \end{array}$$

在图 12.21 中初始表是由其基解表示的，即顶点[0，0]. 第 2 个表由顶点[0，2]表示. 于是，旋转是从顶点沿边 $x=0$ 到相邻顶点[0，2]的一个移动.

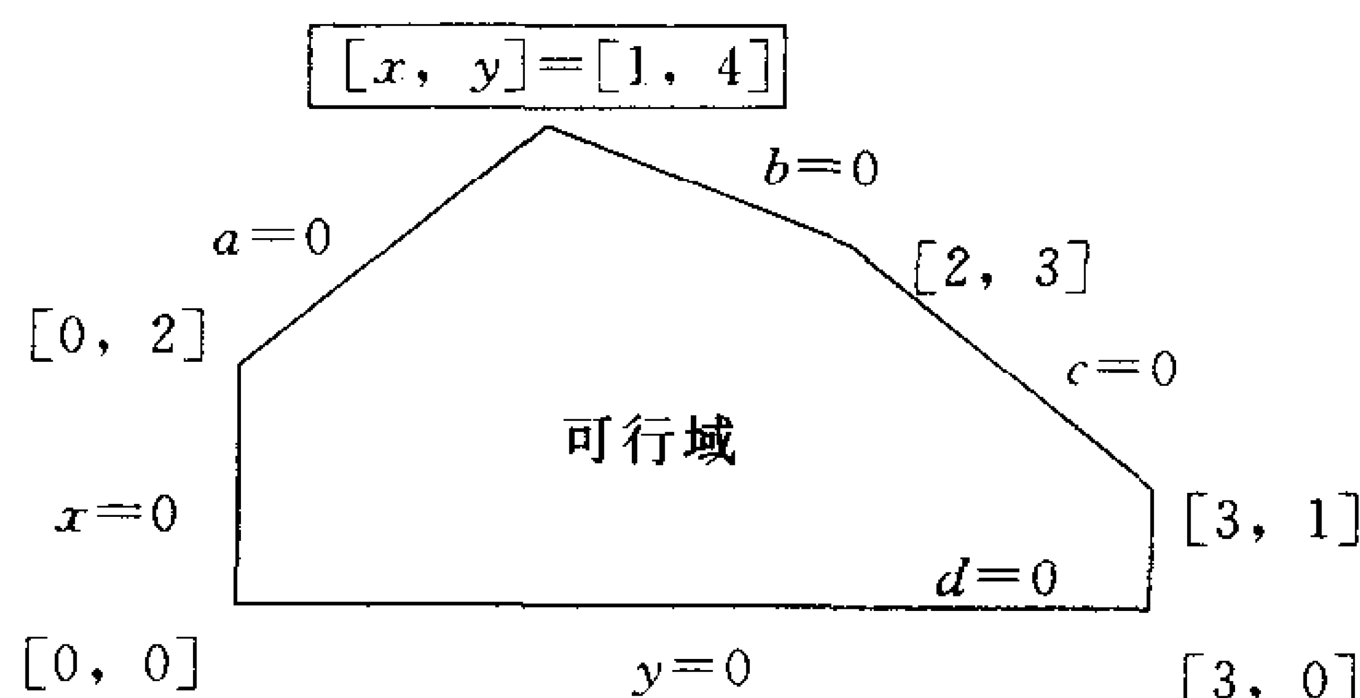


图 12.21　例 12.20 的可行域

现在我们必须选 x 列为旋转列，b 行为旋转行，旋转元是 -3，旋转得到最优表

$$\begin{array}{ccc} b & a & 1 \end{array}$$
$$\begin{bmatrix} * & * & 4 \\ * & * & 1 \\ -4 & 1 & 3 \\ -1 & 0 & 2 \\ 17/3 & 7/3 & -33 \end{bmatrix} \begin{array}{l} =y \\ =x \\ =c \\ =d \\ =-f\to\min \end{array}$$

因为此表是最优的，我们不需要对元素标上 *. 在 $x=1$，$y=4$ 取得 $\min=-33$，$\max=33$. 在图 12.21 中我们从顶点[0，2]沿边 $a=0$ 转到邻接顶点[1，4].

130

如果在初始表中选其他的列为旋转列将会怎样呢? 我们沿着路径[0，0]，[3，0]，[3，1]，[2，3]，[1，4]经过 4 步旋转得到最优解而不是 2 步旋转. ■

作为总结，这里引用 Dantzig 的一段话："单纯形法的巨大功能不断地使我感到惊讶."

练习

1～2. 证明图 12.3 中的菱形和区间是凸集.

3. 证明约束 $x^4+y^4\leqslant 1$ 的可行域是凸的. 这个区域是一个线性规划的可行域吗?

4～6. 证明图 12.4 中的三个集合是非凸的.

7. 证明约束 $|x|\geqslant 1$ 的可行域是非凸的. 由此得出结论：这个区域不是一个线性规划的可行域.

8. 用一个有无穷多个线性约束的系统表示圆盘 $x^2+y^2\leqslant 1$.

9. 开区间 $0<x<1$ 能由线性约束系统给出吗?

10. 找出 5 维单纯形 $x_1+x_2+x_3+x_4+x_5+x_6=1$，所有 $x_i\geqslant 0$ 的顶点.

11. 考虑两列(一个变量在顶端)m 行($m-1$ 个基变量在边上)的可行表．用单纯形法，最大可能的旋转步数是多少？提示：从很小的 m 开始考虑．

12. 考虑三列(两个变量在顶端)m 行($m-1$ 个基变量在边上)的可行表．用单纯形法的第 2 阶段，最大可能的旋转步数是多少？如果能对第 1 次旋转猜到一个更好的旋转元，结果又如何？提示：从很小的 m 考虑并假设旋转步是非退化的，用图解法．

13. 设 S 是 n 元实向量$[x_1, \cdots, x_n]$组成的集合．例如 S 可能是决策变量为 x_i 的线性规划的可行集．证明下面有关 S 的 5 个性质是等价的： 131

(a)S 上的仿射函数是有界的．

(b)函数 x_i，$i=1, \cdots, n$ 在 S 上是有界的．

(c)函数 $x_1^2+\cdots+x_n^2$ 在 S 上是有界的．

(d)函数 $|x_1|+\cdots+|x_n|$ 在 S 上是有界的．

(e)函数$\max\limits_{1\leqslant i\leqslant n}|x_i|$ 在 S 上是有界的．

注 集合 S 称为有界的，是指这些条件之一成立(因此全部成立)．在一些出版物中，有界性包含在多面体或多胞形的概念里．一个有界多面体是它顶点的凸包(顶点的所有凸组合组成的集合)．反之，任意有限集的凸包 S 是一个有界的凸多面体，故 S 可写成线性约束的有限集合．函数 f 在集合 X 上是有界的，是指存在数 C 使得对所有 x 属于 S 都有 $f(x)\leqslant C$.

14. 设 S 是由线性约束系统给出的凸集．证明点 x 是顶点当且仅当没有其他的点与其有同样的紧约束．

15. 假设我们在一个凸集 S 上求一个线性型的最小值且只有一个最优解 x．证明 x 是 S 的顶点．提示：令 $x=(y+z)/2$，考虑

$$\min = f(x) = (f(y)+f(z))/2.$$

16. 对任意线性规划及其可行域 S，证明：对每个 S 的顶点 x 存在一个线性型 f，使得当求 f 在 S 上的最小时，x 是唯一的最优解．构造一个无此线性型 f 的凸集 S 和顶点 x．

17. 令 x 是凸集 S 的极点．证明：存在一个非常数线性型 f，使得 x 是在 S 上极大化 f 的最优解．换句话说，过 x 点存在 S 的正切超平面．提示：要想一般地求解此问题，需要具有博士水平．设法对一些你知道的简单的凸集(例如平面上的凸多边形)解决此问题．

18. 令 x 是凸集 S 的极点，S'是包含 x 的另一个凸集，则 x 是交集 $S\cap S'$的顶点．例如，$x=0$，S 是区间 $0\leqslant x\leqslant 1$，S'是射线 $x\leqslant 1/2$. 132

第5章 对 偶 性

13. 对偶问题

我们回顾一下线性规划问题的典范型：

$$cx + d \to \min, Ax \leqslant b, x \geqslant 0.$$

这里 x 是未知变量组成的列向量，行向量 c，数 d，矩阵 A 和列向量 b 是已知数据．详细地说，我们要求

最小化目标函数

$$f(x) = c_1 x_1 + c_2 x_2 + \cdots + c_m x_m.$$

受限制于线性约束

$$a_{11} x_1 + a_{12} x_2 + \cdots + a_{1m} x_m \leqslant b_1$$

$$a_{21} x_1 + a_{22} x_2 + \cdots + a_{2m} x_m \leqslant b_2$$

$$\vdots$$

$$a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nm} x_m \leqslant b_n$$

和非负约束

$$x_i \geqslant 0, i = 1, 2, \cdots, m.$$

通过引入松弛变量 $u_j \geqslant 0$，$j=1$，2，…，n，前 n 个线性约束可写成线性等式：

$$u_j = b_j - a_{j1} x_1 - a_{j2} x_2 - \cdots - a_{jm} x_m, j = 1, \cdots, n.$$

(即矩阵形式为 $u=b-Ax\geqslant 0.$)

因此，线性规划问题的矩阵形式为

$$\begin{gathered}\text{minimize } cx\\ \text{subject to } b - Ax = u\\ x \geqslant 0, u \geqslant 0.\end{gathered}$$

或者简化为

$$cx \to \min, b - Ax = u \geqslant 0, x \geqslant 0,$$

其中 c 是 $1\times m$ 行矩阵 $c=[c_1\ c_2 \cdots\ c_m]$，x 是 $m\times 1$ 列矩阵

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix},$$

b 是 $n\times 1$ 列矩阵

$$b=\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix},$$

u 是 $n\times1$ 列矩阵

$$u=\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix},$$

且 A 是 $n\times m$ 矩阵

$$A=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1m}\\a_{21}&a_{22}&\cdots&a_{2m}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nm}\end{bmatrix}.$$

可将上述线性规划问题写成标准行表

134

$$\begin{array}{cc} x^{\mathrm{T}} & 1 \end{array}$$
$$\begin{bmatrix}A & b\\ c & d\end{bmatrix}\begin{array}{l}=u\\=f\to\min, x\geqslant 0, u\geqslant 0\end{array}$$

也可以将此问题写成一个列表

$$\begin{array}{r}-x\\1\end{array}\begin{bmatrix}-A^{\mathrm{T}} & c^{\mathrm{T}}\\ b^{\mathrm{T}} & -d\end{bmatrix}\quad x\geqslant 0, u\geqslant 0$$
$$\begin{array}{cc}\| & \|\\ u^{\mathrm{T}} & -f\end{array}\quad\to\max$$

我们称它是标准的. 注意矩阵

$$\begin{bmatrix}-A^{\mathrm{T}} & c^{\mathrm{T}}\\ b^{\mathrm{T}} & -d\end{bmatrix}$$

是一个任意给定的矩阵，所有变量各不相同. 为遵守相同的旋转法则(8.6)，用 $-x$代替 x 得到

$$\begin{array}{r}-x\\-y\end{array}\begin{bmatrix}\alpha^* & \beta\\ \gamma & \delta\end{bmatrix}\mapsto\begin{array}{r}-u\\-y\end{array}\begin{bmatrix}1/\alpha & -\beta/\alpha\\ \gamma/\alpha & \delta-\beta\gamma/\alpha\end{bmatrix}$$
$$\begin{array}{cc}\| & \|\\ u & v\end{array}\qquad\qquad\begin{array}{cc}\| & \|\\ x & v\end{array}$$

于是我们用标准行表的旋转法则来旋转标准列表，旋转元一直是矩阵的非零元，不必在最后一行或最后一列进行选取以保持表是标准的. 我们改变了出基变量的符号，这是与(8.6)的唯一不同之处. 除了边缘外，矩阵是相同的.

定义 下面形式的标准列表的基解

$$\begin{matrix} -y \\ 1 \end{matrix}\begin{bmatrix} A & b \\ c & d \end{bmatrix} \quad y \geqslant 0, v \geqslant 0$$
$$\begin{matrix} \| & \downarrow \\ v & \max \end{matrix}$$

是$y=0$，$v=c$.

这个基解是可行的当且仅当$c \geqslant 0$. 此时，列表称为(列)可行的，相应的目标函数值是d.

定义 标准列表称为最优的，若$b \geqslant 0$，$c \geqslant 0$. 即它既是行可行又是列可行的. 此时，基解$y=0$，$v=c$是最优解且$\min = d$.

135

例 13.1 将饮食问题(例 2.1)写成标准列表

$$\begin{matrix} -a \\ -b \\ -c \\ -d \\ -e \\ 1 \end{matrix}\begin{bmatrix} -0.3 & -73 & -9.6 & 10 \\ -1.2 & -96 & -7 & 15 \\ -0.7 & -20\,253 & -19 & 5 \\ -3.5 & -890 & -57 & 60 \\ -5.5 & -279 & -22 & 8 \\ -50 & -4\,000 & -1\,000 & 0 \end{bmatrix}$$
$$\begin{matrix} =u_1 & =u_2 & =u_3 & =-C \to \max \end{matrix} \quad ■$$

注意到此表是行可行的，故可应用单纯形法的第 2 阶段. 我们将它应用于一个虚构的行问题，后面将其称为对偶问题. 从初始的(原)饮食问题讲，这是对偶单纯形法.

一般地，我们可将任意一个线性规划写成一个标准列表，应用单纯形法——对偶单纯形法. 它比我们在第 4 章讨论的单纯形法更有效.

由标准行表给出的线性规划

$$\begin{matrix} x & 1 \end{matrix}$$
$$\begin{bmatrix} A & b \\ c & d \end{bmatrix} \begin{matrix} = u \\ = z \to \min, x \geqslant 0, u \geqslant 0 \end{matrix} \tag{13.2}$$

称为原始(primal)问题或行问题. 注意标准行表分别用顶端和右边表示行变量 x 和列变量 u. 前面定义的原始问题的对偶(dual)问题或列问题是指下面的线性规划，我们把它写成一个标准列表

$$\begin{matrix} -y \\ 1 \end{matrix}\begin{bmatrix} A & b \\ c & d \end{bmatrix} \quad y \geqslant 0, v \geqslant 0 \tag{13.3}$$
$$\begin{matrix} \| & \| & \\ v & w & \to \max \end{matrix}$$

这里矩阵 y 和 v 的变量完全不同于矩阵 x 和 u 的变量和目标变量 w，矩阵 y 和 v 的变量称为对偶变量.

136

标准列表用余下的边缘(即左端和底端)表示变量．原始对偶线性规划问题可由同一个表给出，所有的边缘都用来标注变量：

$$\begin{array}{rcc}
 & x & 1 \\
-y & A & b \\
1 & c & d \\
 & =v & =w
\end{array}
\quad
\begin{array}{l}
 \\
= u \\
= z \to \min \\
\to \max
\end{array}
\quad
\begin{array}{l}
 \\
x \geqslant 0, u \geqslant 0 \\
y \geqslant 0, v \geqslant 0 \\

\end{array}
\tag{13.4}$$

注意行(原始)问题的每个(原始)变量与列(对偶)问题的每个(对偶)变量是成对的．例如，行矩阵 x 和 v 中的变量是成对的，列矩阵 u 和 y 中的变量是成对的．我们将这种状态称为二元性．二元性的显著特征是(13.2)和(13.3)互为对偶．换言之，原问题(13.2)的对偶问题是问题(13.3)，反之亦然(因此我们可以说“对偶的对偶是原问题”)．这是因为标准行表能重新写成等价的标准列表而且标准列表也可重新写成等价的标准行表，这些结果都是对数据矩阵作相同的运算得到的：

$$\begin{bmatrix} A & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} -A^{\mathrm{T}} & c^{\mathrm{T}} \\ b^{\mathrm{T}} & -d \end{bmatrix},$$

并且**此**运算又重新回到原矩阵(回忆一下对任意矩阵e，$(e^{\mathrm{T}})^{\mathrm{T}}=e$，$-(-e)=e$)．

去掉(13.4)中行问题的基变量 u，得到典范型 $cx^{\mathrm{T}}+d\to\min$，$-Ax^{\mathrm{T}}\leqslant b$，$x\geqslant 0$．其对偶问题的典范型是$b^{\mathrm{T}}y^{\mathrm{T}}-d\to\min$，$A^{\mathrm{T}}y^{\mathrm{T}}\leqslant c^{\mathrm{T}}$，$y\geqslant 0$．因此，回避松弛变量和表，容易将问题的对偶写成典范型．

对于标准型的线性规划$cx^{\mathrm{T}}+d\to\min$，$-Ax^{\mathrm{T}}=b$，$x\geqslant 0$，其对偶可以通过典范型得到．简单对偶问题有与线性方程相应的变量且不要求是非负的．另一个选择是求解标准型的约束系统 $Ax=b$，这让我们得到一个小的典范型和对偶问题．

137

现在我们寻找线性规划和其对偶问题的可行值之间的关系．首先我们有如下定义．

定义 13.5 优化问题的目标函数在其可行域上的值称为可行值．

若可行域是空集(即约束系统不相容)，则无可行值．现在我们来看(13.4)中的一对线性规划．

行问题的可行值是 $cx^{\mathrm{T}}+d$，这里行向量 x 是可行解(即 $x\geqslant 0$，$Ax^{\mathrm{T}}+b=u\geqslant 0$)．列问题的可行值是$-y^{\mathrm{T}}A+d$，其中列向量 y 是可行解(即 $y\geqslant 0$，$-y^{\mathrm{T}}A+c=v\geqslant 0$)．由这些等式可得

$$\begin{aligned} cx^{\mathrm{T}}+d &= (v+y^{\mathrm{T}}A)x^{\mathrm{T}}+d = vx^{\mathrm{T}}+y^{\mathrm{T}}(Ax^{\mathrm{T}})+d \\ &= vx^{T}+y^{T}(u-b)+dvx^{\mathrm{T}}+y^{\mathrm{T}}u-y^{\mathrm{T}}b+d \geqslant -y^{\mathrm{T}}b+d, \end{aligned}$$

因为数vx^{T}，$y^{\mathrm{T}}u\geqslant 0$．

于是我们得到可行值之间的不等式$cx^{\mathrm{T}}+d\leqslant -y^{\mathrm{T}}b+d$和下面的结果．

论据 13.6 最小化线性规划的可行值大于或等于最大化对偶规划的可行值．

这个论据并不违背原始对偶问题之间的对称性，因为在变换时乘了-1. 只要问题可行，这个论据的简短表示是min≥max. 为方便起见，min 和 max 也包含无界和不可行的意思，不等式min≥max对所有情况都成立．也就是说，行问题无界其对偶问题无可行值(即对偶问题不可行，因此max＝min＝$-\infty$). 若列问题无界，则行问题不可行，因此max＝min＝∞.

显然不等式中等号成立当且仅当$vx^{T}+y^{T}u=0$(即取值非 0 的变量其对偶变量取值为 0). 此条件叫做*互补松弛性*(complementary slackness). 我们得到下面的结果，它在证明解是最优时非常重要．

138

论据 13.7 假设线性规划的原始问题和对偶问题都有一个可行解且可行值相等(即互补松弛性成立)，则它们是最优解．

对偶定理(定理 13.8)是一个更进一步的结果，当原始问题和对偶问题都可行时，它用等式max＝min代替min≥max. 特别地，它给出了最后论据的反面(定理 13.9). 对偶定理还叙述了线性规划的原始问题及其对偶问题的有关性质，例如可行性和有界性．其思想是将旋转法应用到线性规划的原始问题和对偶问题，期望得到共同的结论．根据定理 11.3，这些问题都有三个可能的结果．然而，就像我们要看到的一样，原始问题和对偶问题的结果密切相关．

为了观察定理的实施，我们将原始问题和对偶问题写成一个表(如同(13.4))并运用单纯形法．则在有限次旋转内，可能出现下面三个结果．

情形 1：得到一个最优表($b\geq 0$，$c\geq 0$).

情形 2：得到一个有一个坏行的表.

情形 3：得到一个有一个坏列的行可行表($b\geq 0$).

在情形 1，最优表($b\geq 0$，$c\geq 0$)给出两个问题的最优解．原始问题的最优解是$u=b$，$x=0$；对偶问题的最优解是$v=c$，$y=0$. 而且 z 和 w 的最优值都是$d=\min(z)=\max(w)$.

在情形 2，$\min(z)=-\infty$且列问题无可行解．故$\max(w)=-\infty$.

在情形 3，行问题无可行解(即$\min(z)=+\infty$). 为了观察列问题发生的变化，我们转置表格并应用单纯形法．此时，原行问题变成列问题且原列问题现在变成行问题．注意我们得不到最优表，但是它在新表里给出列问题(即原行问题)的最优解．因此只有两个可能的结果：

139

一个有一个坏行的表或者一个有一个坏列的可行表．

若是前者，两个问题都无可行解；若是后者，新的行问题(原列问题)无界．可以证明继续旋转，我们能得到一个有一个坏行和一个坏列的表．

于是我们得到线性规划的主要定理——四择一定理(包括对偶定理).

定理 13.8 对于线性规划及其对偶(13.4)，有且仅有下面一种情形发生：

(1)两个问题都有最优解且最优值相同($\min(z)=\max(w)$).

(2)两个问题不可行．

(3)行问题不可行且列问题无界($\min(z)=\infty=\max(w)$).

(4)列问题不可行且行问题无界($\min(z)=-\infty=\max(w)$).

证明　由避免循环的单纯形法的存在性及前面的讨论可得定理. ■

有时称此定理为三择一定理，因为(3)和(4)是对称的. 但是，通常三择一定理减少了精确分类，即只有情形 1～3. 我们偏爱"四择一定理"的名字.

对偶定理通常是指情形(1)，(2)，(4)关于最优值的等式. 但有时简称定理 13.8 为对偶定理.

对偶定理有几种解释和应用. 几何上，对偶定理是说某些凸集可通过超平面把它和外面的点分离. 这是证明对偶定理的传统方法. 但我们给出另一种证明方法：根据存在有效的单纯形法(避免循环的方法)，可以得到这个结果. 在矩阵博弈理论中(参见第 7 章)，我们将看到对偶定理的另一种解释. 对经济中的问题而言，对偶问题和对偶定理也有重要的经济解释(见下一节的例子).

140

作为对偶定理的推论，我们得到下面的结果，称为*互补松弛定理*.

定理 13.9　*原始问题和对偶问题的可行解是其最优解的充要条件是可行值相同(即互补松弛条件成立).*

问题 13.10　检验 $x_1=1$，$x_2=2$，$x_3=3$，$x_4=4$，$x_5=0$ 是否是下面线性规划的最优解：

$$x_1-x_2+x_3-2x_4+6x_5\to\min,$$

$$x_1+x_2+x_3+x_4+x_5\geqslant 10,$$

$$x_1+2x_2+3x_3+4x_4+5x_5\geqslant 20,$$

$$x_1-x_2+x_3-x_4+x_5\geqslant -2,$$

$$\text{所有 } x_i\geqslant 0.$$

解　首先检验给出的解 x 是否可行. 我们看到它是可行的且第一、三和最后一个约束等式成立(这样的约束称为*有效的*或*紧的*).

再把问题写成标准行表，松弛变量放在右端，对偶变量 y_i 放在同一个表的左端和底端，这里 x_i，$y_i\geqslant 0$：

$$\begin{array}{cccccccc}
 & x_1 & x_2 & x_3 & x_4 & x_5 & 1 & \\
-y_6 & 1 & 1 & 1 & 1 & 1 & -10 & =x_6\\
-y_7 & 1 & 2 & 3 & 4 & 5 & -20 & =x_7\\
-y_8 & 1 & -1 & 1 & -1 & 1 & 2 & =x_8\\
1 & 1 & -1 & 1 & -2 & 6 & 0 & =z\to\min\\
 & =y_1 & =y_2 & =y_3 & =y_4 & =y_5 & =w & \to\max
\end{array}$$

注意在 x 的有效约束中松弛变量 $x_6=x_8=0$. 另一方面，$x_7=10\neq 0$. 假设 x 是最优解，由对偶定理知对偶问题的最优解存在设为 y. 由*互补松弛定理*，$y_1=y_2=y_3=y_4=0$，$y_7=0$. 于是我们的列问题成为

$$\begin{array}{r} -y_6 \\ -y_8 \\ 1 \end{array}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & -10 \\ 1 & -1 & 1 & -1 & 1 & 2 \\ 1 & -1 & 1 & -2 & 6 & 0 \end{bmatrix}$$
$$=0 \quad =0 \quad =0 \quad =0 \quad =y_5 \quad =w \to \max$$

第 2 列说明 $y_8-y_6=1$ 而第 4 列说明 $y_8-y_6=2$，矛盾，故给出的解不是最优的． ■ 141

练习

1～2. 写出标准列表：

1. $5x-6y+2z\to\min$，$[x, y, z]\geqslant 0$，$y\leqslant 7$，$x+y\geqslant 3$.

2. min

$0.39a+0.11b+0.18c+0.21d+0.35e+0.44f+0.25g+0.25h+0.23i+0.24j$

subject to

$15a+15b+15c+15d+15e+15f+10g+15h+15i+10j\geqslant 100$,

$25a+25b+25c+6f+10g+25h+10j\geqslant 100$,

$a+25b+25c+25d+30f+25g+25h+25i+25j\geqslant 100$,

$a+25c+25d+30e+25f+25g+25h+25i+25j\geqslant 100$,

$a+25b+25c+25d+30e+25f+25g+25h+25i+25j\geqslant 100$,

$3a+2b+1c+2d+4e+2f+g+2h+3i+3j\geqslant 70$,

$[a, b, c, d, e, f, g, h, i, j]\geqslant 0$

3. 把练习 1 和练习 2 中的线性规划重写成标准行表．
4. 解问题 13.10 中的线性规划．
5. 证明线性规划的可行值集合是直线上的一个凸集．

6～10. 检验下列各解是否是由标准行表给出的线性规划

$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & 1 \end{array}$$
$$\begin{bmatrix} 7 & -2 & -6 & 6 & -1 & 15 \\ -1 & 1 & 1 & -2 & 2 & -4 \\ -1 & 0 & 1 & 0 & -1 & -2 \\ 4 & 0 & -3 & 5 & 3 & 4 \end{bmatrix} \begin{array}{l} =x_6 \\ =x_7 \\ =x_8 \\ =z\to\min \end{array}$$

的最优解？

6. 基解．
7. 所有 $x_i=1$.
8. $x_1=4$，$x_3=7$，$x_5=x_7=1$，其余 $x_i=0$.
9. $x_1=1$，$x_2=2$，$x_3=3$，其余 $x_i=0$.
10. $x_3=3$，$x_2=x_4=x_5=1$，其余 $x_i=0$. 142

14. 灵敏度分析和参数规划

在实际生活中我们经常不知道精确值或这些值会随着时间发生变化．例如，例 2.1(饮食问题)的 RDA 只能模糊估计且它们依赖于国家和年代．灵敏度分析是考察相关数据怎样影响最优解和最优值的．通常分析很小的变化，大的变化由参数规划(parametric programming)研究．本节对此进行讨论．我们已经考虑了数据的某些值不清楚的问题，但所有可能值的问题需要我们解决．这些值叫做参数(parameters).

首先我们说明怎样找到最优值的变化，它是资源限制和需求限制发生微小变化引起的最优表的改变．

假设我们想知道饮食问题(例 2.1)中最小费用的变化，如果

$$\text{蛋白质需求由 } 50 \text{ 变到 } 51; \tag{14.1}$$

三种需求的变化是：

$$\left.\begin{array}{l}\text{从 } 50, 4\,000, 1\,000 \text{ 到 } 50+\varepsilon_1, 4\,000+\varepsilon_2, 1\,000+\varepsilon_3, \\ \text{其中 } |\varepsilon_i| \leqslant 1, i=1,2,3;\end{array}\right\} \tag{14.2}$$

$$5 \text{ 个价格的变化量分别是 } \varepsilon_a, \varepsilon_b, \varepsilon_c, \varepsilon_d, \varepsilon_e \text{ 且它们的绝对值} \leqslant 1; \tag{14.3}$$

$$\text{所有需求的变化如(14.2)，所有价格的变化如(14.3)}. \tag{14.4}$$

现在我们回答这 4 个问题．注意第二个包含第一个，最后一个包含所有问题．

首先将饮食问题写成标准行表：

$$\begin{array}{cccccc} a & b & c & d & e & 1 \\ \end{array}$$
$$\begin{bmatrix} 0.3 & 1.2 & 0.7 & 3.5 & 5.5 & -50 \\ 73 & 96 & 20\,253 & 890 & 279 & -4\,000 \\ 9.6 & 7 & 19 & 57 & 22 & -1\,000 \\ 8 & 10 & 15 & 5 & 60 & 0 \end{bmatrix} \begin{array}{l} = u_1 \\ = u_2 \\ = u_3 \\ = C \to \min \end{array}$$

[143] (参见例 13.1).

至少需要两次旋转得到最优表：

$$\begin{array}{cccccc} a & b & u_1 & d & u_3 & 1 \\ \end{array}$$
$$\begin{bmatrix} -0.519 & -0.136 & -0.246\,9 & -2.654 & 0.061\,7 & 49.38 \\ -10\,425 & -2\,710 & -4\,941 & -52\,951 & 1\,248 & 996\,931 \\ 0.011 & -0.20 & 0.21 & -0.30 & -0.007\,9 & 2.806 \\ 7.499 & 12.71 & 0.471 & 44.34 & 0.245\,8 & 269 \end{bmatrix} \begin{array}{l} = c \\ = u_2 \\ = e \\ = C \end{array}$$

表中数是近似值，精确的最优值是 $80\,000/297 \approx 269.36$，不是 269. 将表示蛋白质和钙的剩余变量 u_1，u_3 放在顶端，于是相应的约束是有效的(对基解等式成立的约束)而关于维生素 A 的约束不是有效的(有剩余). 因此对维生素 A 的需求量 4 000 的一个小改变不会影响最优解．

要想知道 50 被另一个数代替后最小费用的变化，方法之一是在原表中用参

数 p 代替 50 并旋转这个修正表．此参数将呆在最后一列．然而，还有另一种方法不需要旋转就可以得到修正的最优表．不是用参数 p 代替 50 而是用条件 $u_1 \geq p-50$ 代替松弛变量的限制 $u_1 \geq 0$，此时我们的表不是标准的，但我们已经不需要做旋转了．

现在回答(14.1)．u_1 的约束是 $u_1 \geq 1$，故我们不能把它和顶端的其他变量一起设置为 0 而得到最优解．我们能做的是令 $u_1 = p - 50 = 1$，它给出最优值 $\approx 0.47 + 269.36 = 269.83$. 重要的是验证可行性：$c$ 和 u_2 的值下降但仍为正．因此(14.1)的答案是最小费用大约增加 0.47(精确数是 140/297)．

类似地，如果把需求量 50 用 $p=48$ 代替，我们将条件 $u_1 \geq 0$ 松弛为 $u_1 \geq -2$，可取 $u_1=-2$ 使 C 改善为 280/297(现在最小费用下降)．同样重要的是验证可行性：e 的值下降但仍为正．

于是，当需求量 50 用接近于 50 的参数 p 代替时，我们看到最小费用 $\mathrm{mc}=\mathrm{mc}(p)$ 是一个关于 p 的仿射函数(因此，当松弛变量 $u_1=0$ 变为 $u_1=p-50$ 时，$p=50$ 的最优表生成一个可行解)且数 $140/297 \approx 0.471$ 是 $\mathrm{mc}(t)$ 的斜率．这就是饮食问题中著名的蛋白质的影子价格(shadow price)．我们可以将它和其他来源的蛋白质的价格进行比较，制定饮食策略．

144

我们还可以得到下面的结论．如果我们分别用参数 t_1，t_2，t_3 代替需求限制 50，4 000，1 000，则对充分接近于 0 的 $|t_1-50|$，$|t_2-4\,000|$，$|t_3-1\,000|$，最小费用 $\mathrm{mc} \approx 0.471t_1 + 0.240t_3 + 269.36$ 是 t_1，t_3 的仿射函数且与 t_2 无关．这并不奇怪，因为相应的约束对基解不是有效的——维生素 A 有剩余．因此(14.2)的答案是约等于 $0.471\varepsilon_1 + 0.240\varepsilon_3$.

而且，修正问题的(标准)最优表的最后一列可以写成

$$\begin{array}{c} 1 \\ \begin{bmatrix} 49.38 - 0.246\,9\varepsilon_1 + 0.061\,7\varepsilon_3 \\ 996\,931 - 4\,941\varepsilon_1 + 1\,248\varepsilon_3 + \varepsilon_2 \\ 2.806 + 0.21\varepsilon_1 - 0.007\,9\varepsilon_3 \\ 269 + 0.471\varepsilon_1 + 0.245\,8\varepsilon_3 \end{bmatrix} \end{array} \begin{array}{l} \\ = c \\ = u_2 \\ = e \\ = C \to \min \end{array} \qquad (14.5)$$

该表的其余部分不动，与 ε_i 无关．现在容易计算对哪个 ε_i，该表保持最优．显然当 $|\varepsilon_i| \leq 1$ 时最后一列是正的．

下面我们讨论问题(14.3)．因为对偶问题有相同的最优值且行问题的目标函数的系数是对偶问题的约束的常数项，我们可用对偶性得到答案．我们也可以直接说数据的小变化只能引起最优表的小变化，故表仍是最优的．因此对充分接近于 0 的 ε_* 答案近似为 $48.38\varepsilon_c + 2.806\varepsilon_e$. 但是(14.3)中的数 1 充分接近于 0 吗?

为了回答此问题写出修正表的最后一行(近似地其他行不依赖参数)：

$$\begin{bmatrix} 7.499 - 0.519\varepsilon_c + 0.011\varepsilon_e + \varepsilon_a \\ 12.71 - 0.136\varepsilon_c - 0.20\varepsilon_e + \varepsilon_b \\ 0.471 - 0.2469\varepsilon_c + 0.21\varepsilon_e \\ 44.34 - 2.654\varepsilon_c - 0.30\varepsilon_e + \varepsilon_d \\ 0.2458 + 0.0617\varepsilon_c - 0.0079\varepsilon_e \\ 269 + 49.38\varepsilon_c + 2.806\varepsilon_e \end{bmatrix}^{\mathrm{T}}. \tag{14.6}$$

[145] 显然当$|\varepsilon_*|\leqslant 1$时最后一行是正的．

最后，当我们在(14.4)的条件下改变需求和价格时，修正最优表的最后一列的前三个元素和(14.5)相同，最后一行的前五个元素和(14.6)相同，且最后一行或列的最后一个元素(近似地)是

$$[\varepsilon_c, \varepsilon_e, 1]\begin{bmatrix} -0.2469 & 0.0617 & 49.38 \\ 0.21 & -0.0079 & 2.806 \\ 0.471 & 0.2458 & 269 \end{bmatrix}\begin{bmatrix} \varepsilon_1 \\ \varepsilon_3 \\ 1 \end{bmatrix}.$$

去掉常数项，我们得到(14.4)的一个近似答案．

上述方法可应用于一般问题．也就是说，假设对线性规划 $cx \to \min$，$Ax \geqslant t$，$x \geqslant 0$，考虑需求 t 接近某一值 $t=b$ 时最优值的依赖程度．假设此规划对 $t=b$ 有一个最优解且它是所有基变量都取正值的一个最优表的基解(为将问题化为一个标准行表，我们引入剩余变量 $u=Ax-b$)，则最小值 $\mathrm{mv}=\mathrm{mv}(t)$是一个仿射函数 $c_1(t-b)+\bar{d}$，$\bar{d}=\mathrm{mv}(b)$是最优表的最后一行的最后一个元素，行向量 c_1 由最优表的最后一行对应有效约束的元素和 0 组成，而这些 0 是对应非有效约束的参数值．当目标函数是费用时，行向量 c_1 的元素称为对应有效约束的影子价格．注意非有效约束的需求的小变化不改变最优值．对应的影子价格是 0. 当基松弛变量取 0 时，此约束可能有两个不同的影子价格：一个对应于上升的需求，另一个对应于下降的需求(见第 14 节参数规划的讨论)．

于是对偶问题告诉我们最优值是怎样依赖于数据的某些变化的．类似地，最优表告诉我们最优值是怎样依赖于目标函数的微小变化的，而且这种依赖是仿射的．

注 更一般地，我们用参数 t 的函数 $F(t)$和 $G(t)$(不一定是仿射的)代替 b 和 c，这使得最优值是一个 t 的函数 $\mathrm{mv}(t)$(不一定对所有的 t 有定义)，那么问题$b=F(t_0)$(若存在)的最优表的最后行和列允许我们将 $\mathrm{mv}(t)$的偏导数表示成 $F(t)$和 $G(t)$ 偏导数(若存在)的线性函数． [146]

通过最优表，也可以看出初始表的所有元素(不仅是最后行和列)的微小改变引起的最优值的改变．为简单起见假设仅有一个参数 t，设置如下．我们有一个标准行表

$$\begin{array}{c} x \quad\quad 1 \\ \begin{bmatrix} A(t) & b(t) \\ c(t) & d(t) \end{bmatrix} \begin{array}{l} = u \\ \to \min \end{array} \end{array} \tag{14.7}$$

它与参数 t 有关且所有函数在 $t=t_0$ 可微．$t=t_0$ 时的最优表

$$\begin{array}{c} \bar{x} \quad\quad 1 \\ \begin{bmatrix} \bar{A} & \bar{b} \\ \bar{c} & \bar{d} \end{bmatrix} \begin{array}{l} = \bar{u} \\ \to \min \end{array} \end{array} \tag{14.8}$$

$\bar{b}$，$\bar{c}$ 的元素是正的．我们证明此线性规划对每个接近于 $t=t_0$ 的 t 确实有一个最优解 $x=x(t)$，而且我们算出最优值 $z(t)=c(t)x(t)^{\mathrm{T}}$ 在 $t=t_0$ 的导数 $z'(t_0)$.

为简化记号，我们将表中的行和列转置为 $x=[x_1, x_2]$，$u=\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$，$\bar{x}=[u_1, x_2]$，$\bar{u}=\begin{bmatrix} x_1^{\mathrm{T}} \\ u_2 \end{bmatrix}$. 行向量 x_1 由最优表中即将转入右边的顶端变量组成．它可能是空的，行向量 x_2 由顶端其余变量组成也可能是空的．此时答案是平凡的：$z'(t_0)=d'(t_0)$.

将初始表(14.7)和最优表(14.8)重新改写为

$$\begin{array}{c} x_1 \quad\quad x_2 \quad\quad 1 \\ \begin{bmatrix} \alpha(t) & \beta(t) & b_1(t) \\ \gamma(t) & \delta(t) & b_2(t) \\ c_1(t) & c_2(t) & d(t) \end{bmatrix} \begin{array}{l} = u_1 \\ = u_2 \\ \to \min \end{array} \end{array} \tag{14.9}$$

$$\begin{array}{c} u_1^{\mathrm{T}} \quad x_2 \quad 1 \\ \begin{bmatrix} * & * & \bar{b}_1 \\ * & * & * \\ \bar{c}_1 & * & \bar{d} \end{bmatrix} \begin{array}{l} = x_1^{\mathrm{T}} \\ = u_2 \\ \to \min \end{array} \end{array} \tag{14.10}$$

147

$t=t_0$ 时的表(14.9)转为表(14.10)的旋转将参数表(14.9)变为

$$\begin{array}{c} x_1 \quad\quad\quad x_2 \quad\quad\quad\quad 1 \\ \begin{bmatrix} \alpha(t)^{-1} & -\alpha^{-1}\beta(t) & -\alpha(t)^{-1}b_1(t) \\ \gamma(t)\alpha(t)^{-1} & * & b_2(t)-\gamma(t)\alpha(t)^{-1}b_1(t) \\ c_1(t)\alpha(t)^{-1} & * & d(t)-c_1(t)\alpha(t)^{-1}b_1(t) \end{bmatrix} \begin{array}{l} = u_1 \\ = u_2 \\ \to \min \end{array} \end{array} \tag{14.11}$$

(见注 8.11)．因为表(14.9)中的所有函数在 $t=t_0$ 连续，则对充分接近于 t_0 的 t 我们有：矩阵 $\alpha(t)$ 可逆(因为 $\alpha(t_0)$ 可逆)；表(14.10)的最后一列除最后一个元素外是严格正的(因为 $\bar{b}>0$)；表(14.10)的最后一行除最后一个元素外是严格正的(因为 $\bar{c}>0$)．因此表(14.11)是最优的且它的基解是仅有的最优解(对 t 靠近 t_0)，最后一行的最后一个元素是最优值．它在 $t=t_0$ 的导数是

$$z'(t_0) = d'(t) - c'_1(t_0)\bar{b}_1 - \bar{c}_1 b'_1(t_0) + \bar{c}_1\alpha'(t_0)\bar{b}_1.$$

当 $A(t)$ 是常数时最后一项去掉．当 $A(t)$，$c(t)$ 是常数且是仿射函数或者 $A(t)$，

$b(t)$是常数而 $c(t)$，$d(t)$是仿射函数时最优值是 t(t 靠近 t_0)的仿射函数．

当最优表的某些基变量取 0(这样的表叫做退化的)会出现什么结果？此时斜率可能依赖于参数改变的趋势．

例 14.12 本例取自 Dorfman，Samuelson，Solow (McGraw-Hill，1958) 的书《*Linear programming and economic analysis*》的 4～10 节．

一家化学公司用两种大型的设备(蒸馏器和曲颈瓶)生产一种原材料，公司有四种不同的加工工序可用．若工序 1 用于生产 100 吨原材料，它将利用蒸馏器 7%的周产量和曲颈瓶 3%的周产量．产值和费用与使用的工序有关．若 100 吨由工序 1 生产，公司的净利润是 60 美元．公司计划每周加工 1500 吨原材料．显然公司希望利用 4 道工序的最优组合来获取最大的效益．4 道工序的相关信息如

148

下表：

工　序	(1)	(2)	(3)	(4)	可用
原材料(吨/周)	100	100	100	100	1 500
蒸馏器容量(%)	7	5	3	2	100
曲颈瓶容量(%)	3	5	10	15	100
利润(美元/周)	60	60	90	90	

将这些数据写成行和列标准表，其中 x_i 是工序 i 的水平(强度)($x_1=1$ 表示 100 吨原材料由工序 1 生产)，y_i 是一个松弛变量．此线性规划的标准列表如下：

$$\begin{array}{r} -x_1 \\ -x_2 \\ -x_3 \\ -x_4 \\ 1 \\ \end{array}\begin{bmatrix} 100 & 7 & 3 & -60 \\ 100 & 5 & 5 & -60 \\ 100 & 3 & 10 & -90 \\ 100 & 2 & 15 & -90 \\ 1\,500 & 100 & 100 & 0 \end{bmatrix}\quad \begin{array}{l} x_1, x_2 \geqslant 0, \\ x_3, x_4 \geqslant 0; \\ y_1, y_2, y_3 \geqslant 0 \end{array}$$
$$\qquad =y_1 \quad =y_2 \quad =y_3 \quad =f \quad \to \max$$

我们也可以得到标准行表，对其进行旋转：

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} -100 & -100 & -100 & -100 & 1\,500 \\ -7 & -5 & -3 & -2 & 100 \\ -3 & -5 & -10 & -15 & 100 \\ -60 & -60 & -90 & -90 & 0 \end{bmatrix} \begin{array}{l} =y_1 \\ =y_2 \\ =y_3 \\ =-f\to\min \end{array}$$

我们在行表上进行运算．它是(行)可行的，因为最后一列的前三个元素 1 500，100，100 是正的．因此转到第 2 阶段，进行两次旋转，得到最优表

$$
\begin{array}{ccccc|l}
y_1 & x_2 & y_3 & x_4 & 1 & \\
-1/70 & -5/7 & 1/7 & 5/7 & 50/7 & = x_1 \\
61/700 & 6/7 & -4/7 & -13/7 & 185/7 & = y_2 \\
3/700 & -2/7 & -1/7 & -12/7 & 55/7 & = x_3 \\
33/70 & 60/7 & 30/7 & 150/7 & -7\,950/7 & = -f \to \min
\end{array}
$$

[149]

因此(唯一)答案是每周最大利润 $f=7\,950/7$ 美元(近似为 1 135.71 美元)在 $x_1=50/7$，$x_2=0$，$x_3=55/7$，$x_4=0$ 达到．我们充分利用了可用的原材料($y_1=0$)和曲颈瓶($y_3=0$)，但蒸馏器未充分利用．

购买和使用更多的原材料能得到更多的利润吗？答案依赖于购买额外材料的价格和影子价格．影子价格是指用 1 501 或 1 499 代替 1 500 得到的线性规划最优值的变化量．我们考虑用 $1\,500+\varepsilon$ 代替 1 500 得到的线性规划．这时扰动问题的标准行表为

$$
\begin{array}{ccccc|l}
x_1 & x_2 & x_3 & x_4 & 1 & \\
-100 & -100 & -100 & -100 & 1\,500+\varepsilon & = y_1 \\
-7 & -5 & -3 & -2 & 100 & = y_2 \\
-3 & -5 & -10 & -15 & 100 & = y_3 \\
-60 & -60 & -90 & -90 & 0 & = -f \to \min
\end{array}
$$

旋转两次，得到最优表

$$
\begin{array}{ccccc|l}
y_1 & x_2 & y_3 & x_4 & 1 & \\
-1/70 & -5/7 & 1/7 & 5/7 & 50/7+\varepsilon/70 & = x_1 \\
61/700 & 6/7 & -4/7 & -13/7 & 185/7-61\varepsilon/700 & = y_2 \\
3/700 & -2/7 & -1/7 & -12/7 & 55/7-3\varepsilon/700 & = x_3 \\
33/70 & 60/7 & 30/7 & 150/7 & -7\,950/7-33\varepsilon/70 & \to \min
\end{array}
$$

如果不用旋转，对很小的 ε 怎样得到答案？下面我们回答这个问题．松弛条件 $y_1\geqslant 0$ 变为 $y_1\geqslant -e$. 由最优表我们得到对小正数 ε 扰动松弛问题 f 的最优值是 $7\,950/7+33\varepsilon/70$. ■

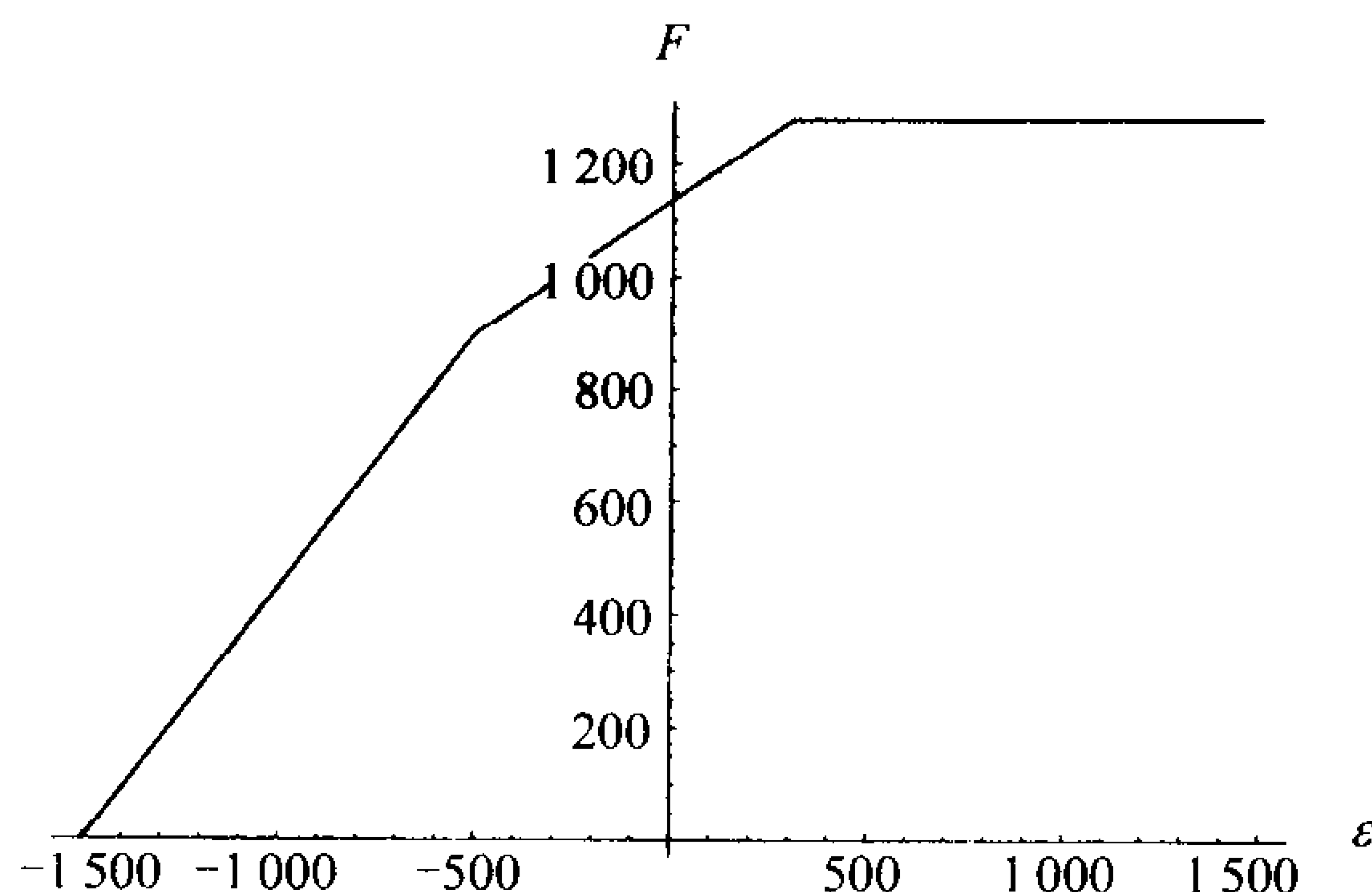
注意扰动问题的最优表的最后一列是原问题最后一列的一个扰动；也就是，它们的差别由一个列矩阵乘以 ε 给出，而且这个列矩阵正是原问题第一列的负值．

问题的答案取决于影子价格(即扰动问题的最优表中 ε 的系数 33/70，也是最优值作为参数函数的斜率)．如果原材料的价格低于影子价格，则购买更多的原材料提高利润，否则不再购买原材料．

[150]

斜率的经济解释与形势有关，可以是影子价格(即我们愿意付给少量附加资源的最大价格)或者边际费用(即生产少量附加产品所需的附加费用，使总费用最小)．

现在我们考虑前面的问题，用 $1\,500+\varepsilon$ 代替 1 500，这里 ε 可取全部实数（在我们只对很小的 ε 感兴趣之前），这是参数线性规划的例子．注意 ε 不是我们控制下的变量，而是一个参数．最优值（我们问题中的最大利润）是参数的函数 $F(\varepsilon)$．当 $\varepsilon\leqslant-1\,500$，无可行解．为了得到利润，我们需要原材料，故 $F(-1\,500)=0$．我们已经求得 $F(0)=7\,950/7$．下面是 $F(\varepsilon)$ 的曲线：



函数 $F(\varepsilon)$ 由线段组成（这种函数经常称为分段线性函数，见下面的定义）．函数非降因为我们不要求所有的资源供应．函数的斜率是原材料的边际费用，它在断点（break point）发生变化．斜率下降（这样的函数叫做凹函数，见下面的定义 14.14）——收益递减律（the law of diminishing return）．

任意线性规划的最优值，只要作为数据参数的函数都有类似的性质．为了描述这些性质我们给出几个定义．

定义 14.13 定义在凸集 P 上的 k 个变量 $t=[t_1,\cdots,t_k]^{\mathrm{T}}$ 的函数 $f(t)$ 称为分段仿射函数，如果集合 P 是有限个凸子集的并集且使得在每个凸子集上 $f(t)$ 是仿射函数．

术语分段线性（piecewise linear）经常用分段仿射（piecewise affine）代替．在最平常情况 $k=1$ 时分段仿射函数的图形由有限个直线段组成．直线上的凸集很容易列出：空集，点，区间，射线，整条直线．

151

注 Hermann Weyl（1885—1955）的话可以帮助我们掌握分段线性函数的概念（思考的数学方法，1940 年在宾夕法尼亚大学 200 周年纪念会议上的一个演讲）：

我们的联邦收入税法规定税 y 是根据收入 x 缴纳的．这种做法以一种很笨的方法将几个线性函数粘贴在一起，它们中的每一个都在收入的另一个区间上是有效的．5 000 年以后的考古学家将会发掘出一些我们的所得税缴款单，与工程遗迹和数学书的碎片同处一处，考古学家推断的时间可能会比现在早几百年，肯定在伽利略和瓦特之前．

定义 14.14 定义在凸集 P 上的 k 个变量 $t=[t_1, \cdots, t_k]^{\mathrm{T}}$ 的函数 $f(t)$ 称为凸函数，如果位于函数 f 图形上面的点的集合是凸的．

有时我们用术语上凹（concave upward）或下凸（convex downward）代替凸．我们称函数 f 是凹的，若函数 $-f$ 是凸的（即位于函数 f 图形下面的点的集合是凸的）．

凸函数有一些等价的定义．例如，（定义在凸集上的）函数是凸的当且仅当它是仿射函数族的最大者．在有限函数族的情况下，最大者是分段仿射凸函数．

现在我们给出参数规划的一般结果．

定理 14.15 考虑线性规划的典范型 $cx\to\min$，$Ax\leqslant b$，$x\geqslant 0$，假设行向量 c 或列向量 b 的所有元素都是 k 个参数 $t_1, \cdots, t_k$ 的函数，P 是使得线性规划有一个最优值的参数 $t=[t_1, \cdots, t_k]^{\mathrm{T}}$ 的值组成的集合，$f(t)$ 是最优值，则 P 是凸集且 $f(t)$ 是 P 上有限个仿射函数的最小者．因此 $f(t)$ 是分段仿射凹函数．若 b 是 t 的非减函数或 c 是 t 的非增函数，则 $f(t)$ 是 t 的非减函数．

证明 如果 P 是空集，结论显然成立．假设我们的规划对 t 的至少一个值有一个最优解．

由对偶定理，只需要考虑参数在目标函数 $c=c'+c''t$ 中的情形，此时可行域 S 与 t 无关．为证明 P 是凸集，考虑 P 中的数 t'，t'' 的凸组合 $t=at'+(1-a)t''$，需证明 $t\in P$，即函数 $c(t)x$ 在 S 上有下界．因 $c(t)$ 是 t 的仿射函数，故 $c(t)=ac(t')+(1-a)c(t'')$．因对数 C 和所有 $x\in S$ 有 $c(t')x$，$c(t'')x\geqslant C$，故对所有 $x\in S$ 有 $c(t)x=ac(t')x+(1-a)c(t'')x\geqslant aC+(1-a)C=C$，因此 $c(t)x$ 在 S 上有下界．

LP 的每个表只在最后一行有参数且最后一行的每个元素都是参数的仿射函数，这对初始表是正确的，并且旋转保持了此性质．这样的表有有限个（它的上界见第 10 节）．

现在考虑对 t 的至少一个值（此值对不同的表可能不同）是最优的表的子集和这些表的最后一行的最后一个元素．对每个 t 这些表是可行的且至少有一个是最优的，因此对这些最后元素而言最优值是最小的．

显然目标函数的所有系数增加或保持不变，$f(t)$ 的最小值不改变． ■

当我们想把几个目标合并成一个目标函数时，带参数的目标函数出现在目标规划里．一种方法是取几个目标的线性组合，我们要做带非负系数的最小化．系数（权重）的选择可能是有争议的，但是它们可以被当作参数．如果我们合并的函数是仿射的，则得到的目标函数也是仿射的，它的系数是参数的仿射函数．因此可以应用定理 14.15．

问题 14.16 求解 $P=3x+4y\to\max$，$2x+2y\leqslant 200$，$x+3y\leqslant t$，x，$y\geqslant 0$，t 是一个给定的数．

解 可作图求解此问题（见第 3 节）．下面是一个提纲．当 $t<0$，可行集是空

152

的，问题不可行．当 $0\leqslant t\leqslant 100$，第 1 个约束是多余的，可行域是一个三角形，当 $x=t$，$y=0$ 时 $\max=3t$. 当 $100\leqslant t\leqslant 300$，最优解是直线 $2x+2y=200$ 和 $x+3y=t$ 的交集．因此当 $x=150-t/2$，$y=t/2-50$ 时 $\max=250+t/2$. 当 $t\geqslant 300$，第 2 个约束是多余的，可行域是一个三角形，当 $x=0$，$y=100$ 时 $\max=400$.

153

现在我们用单纯形法解此问题．引入两个松弛变量 $u=100-x-y\geqslant 0$ 和 $v=t-x-3y\geqslant 0$ 并写成标准行表

$$\begin{array}{c} \begin{matrix} x & y & 1 \end{matrix} \\ \begin{bmatrix} -1 & -1 & 100 \\ -1 & -3 & t \\ -3 & -4 & 0 \end{bmatrix} \end{array} \begin{array}{l} =u \\ =v \\ =-P \quad \to \min \end{array} \tag{14.17}$$

当 $t<0$，v 行是坏列，故问题不可行．现在假设 $t>0$，表是可行的，进入第 2 阶段．我们选取 y 列为旋转列．

如果 $t\geqslant 300$，我们在 -1 旋转得到

$$\begin{array}{c} \begin{matrix} x & u & 1 \end{matrix} \\ \begin{bmatrix} -1 & -1 & 100 \\ 2 & 3 & t-300 \\ 1 & 4 & -400 \end{bmatrix} \end{array} \begin{array}{l} =y \\ =v \\ =-P \quad \to \min \end{array} \tag{14.18}$$

这是最优表，因此在 $x=0$，$y=100$，得到 $\max=400$.

现在假设 $0<t<300$. 根据单纯形法，在 -3 旋转表(14.17)得到

$$\begin{array}{c} \begin{matrix} x & v & 1 \end{matrix} \\ \begin{bmatrix} -2/3 & -1/3 & 100-t/3 \\ -1/3 & -1/3 & t/3 \\ -5/3 & 4/3 & -4t/3 \end{bmatrix} \end{array} \begin{array}{l} =u \\ =y \\ =-P \quad \to \min \end{array} \tag{14.19}$$

现在 x 列是旋转列．比较 $(100-t/3)/(-2/3)$ 和 $(t/3)/(-1/3)$ 选取旋转元．当 $0<t\leqslant 100$ 时在 $-1/3$ 旋转得到

$$\begin{array}{c} \begin{matrix} y & v & 1 \end{matrix} \\ \begin{bmatrix} 2 & 1 & 100-t \\ -3 & -1 & t \\ 5 & 3 & -3t \end{bmatrix} \end{array} \begin{array}{l} =u \\ =x \\ =-P \quad \to \min \end{array} \tag{14.20}$$

因此在 $x=t$，$y=0$，$\max=3t$. 当 $100\leqslant t\leqslant 300$ 时在 $-2/3$ 旋转表(14.19)得到

$$\begin{array}{c} \begin{matrix} u & v & 1 \end{matrix} \\ \begin{bmatrix} -3/2 & 1/2 & 150-t/2 \\ 1/2 & -1 & t/2-50 \\ 5/2 & 1/2 & -t/2-250 \end{bmatrix} \end{array} \begin{array}{l} =x \\ =y \\ =-P \quad \to \min \end{array} \tag{14.21}$$

因此在 $x=150-t/2$，$y=t/2-50$，得到 $\max=t/2+250$. [154]

注意斜率是下降的(收益递减律)，

区间	$0\leqslant t\leqslant 100$	$100\leqslant t\leqslant 300$	$300\leqslant t$
斜率	3	1/2	0

当 $t=200$ 时最后的表是最优的(线性规划的一种情形)，斜率是 v 列的最后元．如果我们开始改变第一个限制 200(而不是第二个限制 200)，斜率将是 5/4. ■

练习

1～4. 求解下列线性规划，所有变量 a，b，c，d，e，f，g，h，i，j 非负．提示：非标准表的行和列问题不一定相互对偶．

1.
$$\begin{array}{c} \\ g \\ h \\ -1 \\ \end{array}\begin{array}{c} \begin{array}{cccc} a & b & c & 1 \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & -1 & -2 \\ 2 & -1 & 0 & -3 \\ 0 & 2 & 1 & 0 \end{array}\right] \\ \begin{array}{cccc} =i & =j & =k & =u \end{array} \end{array}\begin{array}{l} \\ =d \\ =c \\ =w \rightarrow \min \\ \rightarrow \min \end{array}$$

2.
$$\begin{array}{c} \\ g \\ h \\ -1 \\ \end{array}\begin{array}{c} \begin{array}{cccc} a & b & c & 1 \end{array} \\ \left[\begin{array}{cccc} 1 & 2 & 3 & -1 \\ 2 & 0 & 1 & 3 \\ -1 & 1 & 0 & 0 \end{array}\right] \\ \begin{array}{cccc} =i & =j & =k & =u \end{array} \end{array}\begin{array}{l} \\ =d \\ =c \\ =w \rightarrow \min \\ \rightarrow \min \end{array}$$

3.
$$\begin{array}{c} \begin{array}{cccc} a & a & a & -1 \end{array} \\ \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array}\begin{array}{l} \\ =d \\ =e \\ =w \rightarrow \min \end{array}$$

4.
$$\begin{array}{c} \begin{array}{ccc} a & b & -1 \end{array} \\ \left[\begin{array}{ccc} 1 & 0 & 1+\varepsilon \\ 0 & 1 & 1 \\ 1 & 2 & 0 \end{array}\right] \end{array}\begin{array}{l} \\ =c \\ =d \\ =w \rightarrow \max \end{array}$$

其中 ε 是一个给定的数． [155]

15. 对偶性的其他问题

对偶定理有多种解释，是由许多学者提出并以不同形式发表的，其中包括 Fourier(1826)，Gordan(1873)，Minkowski(1896)，Farkas(1901).

定理 6.16 给出的是线性方程组的对偶定理．它的另一种形式是：线性方程

组 $Ax=b$ 有解当且仅当对任意满足 $yA=0$ 的行向量有 $yb=0$ 成立．这是 Fredholm (1903) 提出的．

下面的系统有且只有一个是可行的：

$$(a)Ax=b;\ (b)yA=0,\ yb>0.$$

我们将给出线性不等式系统的对偶定理．首先我们给出由原问题导出对偶问题的方法．主要思想是利用原问题的线性约束找到它的最优值的一个下界，即我们线性地合并给定的约束来得到这个下界．对偶问题转化为：最大化此下界．

我们先看一个简单的例子．假设我们要求解线性规划

$$\begin{cases}\text{minimize} & x+y\\ \text{subject to} & x+y\geqslant 2.\end{cases}$$

显然 $\min(x+y)=2$.

下面是一个复杂的例子：

$$\begin{cases}\text{minimize} & 4x+5y\\ \text{subject to} & x+3y\geqslant 2,\\ & 2x-y\geqslant 3.\end{cases}$$

如果我们将第 1 个约束乘 2 加到第 2 个约束，则得到 $4x+5y\geqslant 7$. 此约束是给定约束的正系数线性组合，它(在我们的条件下)给出目标函数一个下界：$\min(4x+5y)\geqslant 7$. 而且容易看出这个下界能够达到，即它是紧的．于是

156

$$\min(4x+5y)=7.$$

我们试图把这种方法用于较大的 LP，其标准行表如下：

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & 1 \end{array}$$
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5\\ 1 & -2 & 0 & 4 & -1\\ 0 & 2 & -3 & 1 & 0\\ 1 & 2 & 3 & -4 & 5\end{bmatrix}\begin{array}{l} =u_1\geqslant 0\\ =u_2\geqslant 0\\ =u_3\geqslant 0\\ =z\to\min\end{array}\quad \text{所有 } x_i\geqslant 0$$

或写成矩阵形式

$$\begin{array}{cc} x & 1\end{array}$$
$$\begin{bmatrix} A & b\\ c & d\end{bmatrix}\begin{array}{l} =u\geqslant 0\\ =z\to\min\end{array}\quad x\geqslant 0$$

或写成矩阵乘积形式

$$Ax^{\mathrm{T}}\geqslant -b,x\geqslant 0,cx^{\mathrm{T}}+d\to\min.$$

若 y 是一个满足 $y\geqslant 0$，$y^{\mathrm{T}}\cdot A\leqslant c$ 的列矩阵，则我们得到 $cx^{\mathrm{T}}\geqslant u^{\mathrm{T}}\cdot Ax^{\mathrm{T}}\geqslant -y^{\mathrm{T}}\cdot b$. 于是通过对给定的约束 $Ax^{\mathrm{T}}\geqslant -b$ 用非负系数 y 进行线性组合，我们得到目标函数 z 的一个下界

$$z=cx^{\mathrm{T}}+d\geqslant -y^{\mathrm{T}}\cdot b+d.$$

下面我们设法选择如同前面那样的 y 使得这个下界尽可能地紧：

$$-y^{\mathrm{T}}\cdot b+d\rightarrow \max, y\geqslant 0, y^{\mathrm{T}}\cdot A\leqslant c.$$

这是对偶问题，在相同的标准表中它可写成列问题：

$$\begin{array}{rccll} & x & 1 & & \\ -y & \left[\begin{array}{c} A \\ c \end{array}\right. & \left.\begin{array}{c} b \\ d \end{array}\right] & \begin{array}{l} =u \\ =z\rightarrow \min \end{array} & \begin{array}{l} x\geqslant 0, u\geqslant 0 \\ y\geqslant 0, v\geqslant 0 \end{array} \\ 1 & & & & \\ & =v & =w & \rightarrow \max. & \end{array}$$

事实上，矩阵形式便于计算且适应于任意由标准行表给出的 LP. [157]

求 $y\geqslant 0$ 使得 $y^{\mathrm{T}}\cdot A\leqslant c$ 等价于将线性函数 cx 写成所有给定的约束 $Ax\geqslant -b$，$x\geqslant 0$ 带非负系数 y，u 的左边 Ax，x 的线性组合 $y^{\mathrm{T}}Ax+ux$，其中 u 是对偶(列)规划的基变量组成的行矩阵．

对偶定理和我们的问题如何相关呢？一种可能是我们的问题无界．用任何方法我们也不能得到 z 的界，这是显然的．因此对偶问题不可行，这是定理的一部分．另一种可能是我们的问题不可行，由定理知对偶问题无界或不可行．从线性组合来看，这意味着我们或者得到任意好的界或者作为线性组合得不到界．最后一种可能是我们的问题有最优解，则定理说明最优值是由线性组合得到的最好的界．

现在我们问自己一个问题：对哪个数 e 约束 $cx\geqslant d'$ 来自给出的约束 $Ax^{\mathrm{T}}\geqslant -b$，$x\geqslant 0$. 换言之(见第 4 节)，系统的每个可行解都满足此约束吗？显然答案依赖于系统可行域上 $\min(cx)$ 和数 d' 的关系．如果 $\min(cx)<d'$(包括情形 $\min=-\infty$，即规划无界)，则答案是否定的．如果 $\min(cx)\geqslant e$(包括情形 $\min=\infty$，即规划不可行)，则答案仍是否定的．

因此我们从下面的对偶定理的变形中很容易得到我们的变形．

定理 15.1 给定任意不等式系统 $A'x'\geqslant b'$ 和由此产生的另一个不等式 $c'x'\geqslant d'$，则此不等式是此系统中的约束和约束 $0\geqslant -1$ 的非负系数的线性组合，或者此系统是不可行的且约束 $0\geqslant -1$ 是此系统中的约束的一个线性组合．

我们在定理的条件中加上一撇(′)，是因为此时的数据和原来表中的不相同．为了运用定理，我们取 $A'=\begin{bmatrix} A \\ I_n \end{bmatrix}$，$n$ 是 x 中的变量个数，$x'=x$，$b'=-b$，等等． [158]

利用标准技巧 $x=x'-x''$，x'，$x''\geqslant 0$，将数据写在标准行表中，易得定理 15.1.

现在我们叙述定理 6.16 对于不等式系统的变形．假设给定不等式系统 $Ax\geqslant b$ 和同类型的另一个不等式 $cx\geqslant d$，后面的约束是由此系统产生的吗？换句话说(见第 4 节)，系统的每个可行解都满足此约束吗？

显然答案依赖于系统的可行域上的 $\min(cx)$ 和数 d 的关系．如果 $\min(cx)<d$(包括情形 $\min=-\infty$，即规划无界)，则答案是否定的．如果 $\min(cx)\geqslant d$(包括情形 $\min=\infty$，即规划不可行)，则答案是肯定的．

注 15.2 对偶定理意味着任何原始-对偶线性规划

$$cx+d\to\min, Ax\geqslant -b, x\geqslant 0;\quad -yb+d\to\max, yA\geqslant c, y\geqslant 0$$

都可写成一个线性约束系统

$$Ax\geqslant -b, x\geqslant 0, cx+yb=0, y\geqslant 0, yA\leqslant c.$$

此系统的每个可行解$[x^{\mathrm{T}}, y]$给出两规划的最优解 x 和 y，优化问题的最优解 x 和 y 给出系统的一个解$[x^{\mathrm{T}}, y]$. 在此意义下，线性规划就是求解线性约束系统. 寻求 LP 最优解的另一种方法是转化为另一个 LP 的第 1 阶段.

对一个来源于实际生活的线性规划，其对偶规划也可以用生活中的术语来解释. 现在我们考虑例 2.1，2.2，2.3，2.4(更一般的形式)，并给出对偶问题的经济解释.

例 15.3 考虑一般的饮食问题(例 2.1 的推广)：

$$Ax\geqslant b, x\geqslant 0, C=cx\to\min,$$

这里 x 中的 m 个变量代表不同的食物，n 个约束 $Ax\geqslant b$ 代表成分. 在费用 C 最小的情况下利用给定的食物满足对成分的需求 b.

159

另一方面，考虑以价格 y_1，…，$y_n\geqslant 0$ 出售食物成分的大商店. 它的目标是给每种食物匹配价格 $yA\leqslant c$ 使利润 $P=yb$ 最大.

利用松弛变量 $u=Ax-b\geqslant 0$ 和 $v=c-yA\geqslant 0$ 我们将问题写成一个标准表：

$$\begin{array}{cc} & \begin{array}{cc} x^{\mathrm{T}} & 1 \end{array} & & \\ \begin{array}{c} -y^{\mathrm{T}} \\ 1 \end{array} & \left[\begin{array}{cc} A & -b \\ c & 0 \end{array}\right] & \begin{array}{l} =u \\ =C\to\min \end{array} & \begin{array}{l} x\geqslant 0, u\geqslant 0 \\ y\geqslant 0, v\geqslant 0 \end{array} \\ & \begin{array}{cc} =v & =P \end{array} & \to\ \max & \end{array}$$

因此这两个问题互为对偶. 特别地，单纯形法可求解它们，如果它们都可行，则 $\min(C)=\max(P)$. 可以证明第 14 节提到的影子价格是对偶问题中成分的最优价格，故也称*对偶价格*(dual price).

上述数学问题还有另一种经济解释，x 中的变量是不同工序的强度，约束对应着不同的工序，b 是被完成的顺序(或被满意的要求). 使用给定的工序和满足给定的产品限额所需的总费用是 $C=cx$，我们要使它最小. 根据原问题的这个解释，相应有一个对对偶问题的解释：竞争者 Ann 丢失了政府的合同，但她说你可以从她那里以她的低价 $y\geqslant 0$ 买进产品，此价格与你得到的每个工序单位费用($yA\leqslant c$)相匹配且使她的利润 yb 最大.

根据此解释，最优价格 y 对你来说就是*边际费用*(marginal costs)(即生产一个单位的额外产品的额外费用). 从第 14 节我们知道边际费用随着数量(在线性规划中)的增加而下降. ■

例 15.4 考虑一般的混合问题(例 2.2 的推广)

$$Ax=b, x\geqslant 0, C=cx\to\min,$$

这里 x 中的 m 个变量代表不同的合金，n 个约束 $Ax\geqslant b$ 代表元素. 我们需要在

费用 C 最小的情况下利用给定的合金满足对元素的需求 b. 160

另一方面，考虑以价格 y_1，…，y_n 买卖元素的经销商．价格为正表示经销商卖出，价格为负表示经销商买入．经销商的目标是对合金匹配满足 $yA \leqslant c$ 的价格使利润 $P=yb$ 最大．

利用标准技巧和人工变量，将问题写成标准表

$$u' = Ax - b \geqslant 0, u'' = -Ax + b \geqslant 0;$$
$$v = c - yA \geqslant 0; y = y' - y'', y' \geqslant 0, y'' \geqslant 0.$$

■

现在我们将两个问题写进同一个标准表

$$\begin{array}{c} \\ -y'^{\mathrm{T}} \\ -y''^{\mathrm{T}} \\ 1 \\ \end{array} \begin{array}{c} \begin{array}{cc} x^{\mathrm{T}} & 1 \end{array} \\ \begin{bmatrix} A & -b \\ -A & b \\ c & 0 \end{bmatrix} \\ \begin{array}{cc} =v & =P \end{array} \end{array} \begin{array}{l} \\ =u' \\ =u'' \\ =C \to \min \\ \to \max \end{array} \qquad \begin{array}{l} x \geqslant 0, u \geqslant 0 \\ y', y'' \geqslant 0, v \geqslant 0 \end{array}$$

注 15.5 我们将目标函数 C 加上一个任意的常数 d，得到一个任意线性规划 LP 的标准形式，用 d 代替最后一行的最后零元，得到标准行表的 LP.

例 15.6 考虑一般的制造问题(例 2.3 的推广)：

$$P = cx \to \max, Ax \leqslant b, x \geqslant 0,$$

这里 x 中的变量代表产品的数量，P 是想要最大化的利润(或收入)，约束 $Ax \geqslant b$ 代表资源(例如，不同类型的劳力、使用的清水、散发的污染物质、稀有的原材料)，b 代表拥有的资源量．于是对偶问题

$$yb \to \min, yA \geqslant c, y \geqslant 0$$

可以这样解释．你的竞争者 Bob 提议出钱让你放弃以下的条款：你退出买卖，他以你的价格 $y \geqslant 0$ 买进所有的资源，在符合你可生产的每件产品想得的利润下，最小化他的费用． 161

由对偶定理，Bob 的最优价格又成了你资源的影子价格．一种资源的影子价格表明该资源的限制 b_0 没增加一个单位所带来的利润的增加量或减少一个单位带来的利润的减少量．我们只改变 b_0 而不改变其他资源的限制和其他数据．只有有限多个对应着向下的和向上的影子价格的 b_0 值不同．它们其中之一一定是绑定约束或非绑定约束对应的 b_0 值的边界线(在这个意义下去掉该约束不会改变最优值).

从第 14 节我们知道即使资源的供给 b_0 增加资源的影子价格也不增加(收益递减律).

■

例 15.7 本例类似于例 2.4，只是地理位置不同．仓库设在 Bedford 和 Scranton，分别能供应 220 和 280 单位．在 State College、Altoona、Harrisburg 有零售店，分别需要 170、120 和 210 单位．运输费用如下表：

	State College	Altoona	Harrisburg
Bedford	77	39	105
Scranton	150	186	122

约束是

$$x_{ij} \geqslant 0, i = 1,2; j = 1,2,3 \tag{i}$$

$$\begin{cases} x_{11} + x_{12} + x_{13} \leqslant 220 \\ x_{21} + x_{22} + x_{23} \leqslant 280 \end{cases} \tag{ii}$$

$$\begin{cases} x_{11} + x_{21} \geqslant 170 \\ x_{12} + x_{22} \geqslant 120 \\ x_{13} + x_{23} \geqslant 210 \end{cases} \tag{iii}$$

注意此 LP 中仓库能提供的单位总和是 500，等于各零售店所需单位之和，这与前面的例子不同(仓库有 130 件可用的小器具，而零售店只需要 100 个).

由约束(ii)，我们得到

162

$$x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} \leqslant 500. \tag{iv}$$

另一方面，由约束(iii)可得

$$x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} \geqslant 500. \tag{v}$$

合并不等式(iv)和(v)我们得到等式

$$x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} = 500.$$

此等式表明供应量与需求量相等，称之为**平衡条件**. 只要我们愿意，此条件允许我们将(ii)和(iii)中的不等号换成等号. 平衡条件迫使可行解中松弛变量取 0，因此如果我们用等号代替(ii)和(iii)中的所有$\leqslant$，$\geqslant$就可得到一个等价问题.

用单纯形法很容易解决这个小问题. 下一章我们将看到单纯形法特别适合求解运输问题. 现在我们的目的是给出对偶问题的经济解释. 首先我们用例 15.7 给出势的概念. 引入松弛变量将问题化为标准行表，再在同一表中写出对偶变量.

$$\begin{array}{r|ccccccc|l} & x_{11} & x_{12} & x_{13} & x_{21} & x_{22} & x_{23} & 1 & \\ \hline -u_1 & -1 & -1 & -1 & 0 & 0 & 0 & 220 & = y_1 \\ -u_2 & 0 & 0 & 0 & -1 & -1 & -1 & 280 & = y_2 \\ -v_1 & 1 & 0 & 0 & 1 & 0 & 0 & -170 & = z_1 \\ -v_2 & 0 & 1 & 0 & 0 & 1 & 0 & -120 & = z_2 \\ -v_3 & 0 & 0 & 1 & 0 & 0 & 1 & -210 & = z_3 \\ 1 & 77 & 39 & 122 & 150 & 186 & 105 & 0 & = C \\ \hline & \| & \| & \| & \| & \| & \| & \downarrow & \\ & w_{11} & w_{12} & w_{13} & w_{21} & w_{22} & w_{23} & \max & \end{array}$$

$$u_i, \ v_j, \ w_{ij}, \ x_{ij}, \ y_i, \ z_j \geqslant 0$$

对偶问题最大化的目标函数是

$$-220u_1-280u_2+170v_1+120v_2+210v_3. \tag{15.8}$$

对偶问题的控制变量 u_i，v_j 称为势(potentials)．而势对应着每个零售店和仓库的约束(或相应的松弛变量)，对偶问题的其他变量 w_{ij} 对应着原问题的决策变量 x_{ij}.

163

它们是这 6 个对偶约束的松弛变量

$$w_{ij}=c_{ij}+u_i-v_j\geqslant 0\ \forall\, i,j, \tag{15.9}$$

这里 $i=1$，2，$j=1$，2，3，c_{ij} 表示费用矩阵中的元素

$$c=\begin{bmatrix}77 & 39 & 105\\150 & 186 & 122\end{bmatrix}.$$

那么对偶问题的含义是什么?

假设你想成为一个运输商并提出一个简化的价格系统．也就是说，你对每个仓库指定当地价格 $u_i\geqslant 0$，$i=1$，2，对每个零售店指定当地价格 $v_j\geqslant 0$，$j=1$，2，3．为了打败竞争对手，你需要约束(15.9)．你的利润是(15.8)，想要它最大．■

注 15.10 可以看出在下面的变化下这个问题是不变的：

$$u_i\to u_i+t, v_j\to v_j+t,$$

其中对所有 i，j 而言，t 是一个任意的固定值，这允许我们可以忽略约束 u，$v\geqslant 0$.

练习

1～4. 考察系统中最后一个方程是否多余(即可由其他推出).

1. $\begin{cases}x+2y+3z=4\\5x+6y+7z=8\\6x+8y+10z=0\end{cases}$

2. $\begin{cases}x+2y+3z=4\\5x+6y+7z=8\\7x+10y+13z=16\end{cases}$

3. $x=6$，$y=5$，$z=0$，$2x-8y+3z=7$

4. $\begin{cases}x_1+2x_2+3x_3+4x_4+5x_5=6\\6x_1+5x_2+4x_3+3x_4+2x_5=1\\x_1-x_2+x_3-x_4+x_5=0\end{cases}$

5～8. 考察系统中最后一个约束是否多余(即可由其他推出).

$$5.\ \begin{cases} x+2y+3z=4 \\ 5x+6y+7z=8 \\ 6x+8y+10z\geqslant 0 \end{cases}$$

$$6.\ \begin{cases} x+2y+3z\geqslant 4 \\ 5x+6y+7z\geqslant 8 \\ 7x+10y+13z\geqslant 16 \end{cases}$$

164

7. $x=6$, $y=5$, $z=0$, $2x-8y+3z\leqslant 7$

$$8.\ \begin{cases} x_1+2x_2+3x_3+4x_4+5x_5=6 \\ 6x_1+5x_2+4x_3+3x_4+2x_5=1 \\ x_1-x_2+x_3-x_4+x_5\geqslant 0 \end{cases}$$

9～14. 解线性规划，这里所有的 $x_i\geqslant 0$. 提示：用图解法求解对偶问题 .

$$9.\ \begin{array}{cccccccccc|l} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & 1 & \\ \hline 0 & 8 & -5 & 6 & 7 & 8 & 3 & 5 & 4 & -1 & \geqslant 0 \\ -1 & 2 & -2 & 1 & 1 & 1 & 2 & 2 & 5 & 0 & \to \min \end{array}$$

$$10.\ \begin{array}{cccccccccc|l} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & 1 & \\ \hline 6 & 8 & 5 & 6 & 7 & 8 & 3 & 5 & 4 & -1 & \geqslant 0 \\ 1 & 2 & 5 & 1 & 1 & 1 & 5 & 0 & 5 & 0 & \to \min \end{array}$$

$$11.\ \begin{array}{cccccccc|l} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & 1 & \\ \hline 3 & 4 & 1 & 1 & 1 & 2 & 3 & -1 & =x_8 \\ 4 & 3 & 1 & 0 & 0 & 2 & 4 & -2 & =x_9 \\ 5 & 5 & 1.4 & 5 & 6 & 3 & 6 & 0 & =z\ \to \min \end{array}$$

$$12.\ \begin{array}{ccccccccc|l} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & 1 & \\ \hline 6 & 4 & 1 & 10 & 5 & 2 & 8 & 7 & -1 & =x_9 \\ 19 & 9 & 10 & 1 & 9 & 10 & 6 & 7 & -2 & =x_{10} \\ 20 & 10 & 10 & 10 & 10 & 11 & 10 & 10 & 0 & =z\ \to \min \end{array}$$

$$13.\ \begin{array}{cccccccccc|l} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & 1 & \\ \hline 6 & 8 & 5 & 6 & 7 & 8 & 3 & 5 & 4 & -1 & \geqslant 0 \\ 29 & 29 & 14 & 14 & 13 & 13 & 4 & 8 & 3 & -2 & \geqslant 0 \\ 31 & 32 & 15 & 15 & 15 & 15 & 5 & 20 & 5 & 0 & \to \min \end{array}$$

$$
14.\quad \begin{array}{c} \\ \\ \\ \\ \end{array}
\begin{array}{cccccccccc}
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & 1 \\
\end{array}
$$

$$
14.\ \begin{bmatrix} 6 & 8 & 5 & 6 & -7 & 8 & 3 & 5 & 4 & -1 \\ -9 & 0 & -4 & 1 & 13 & 13 & 4 & 8 & 3 & -2 \\ -1 & 2 & -5 & 15 & 15 & 15 & 5 & 0 & 5 & 0 \end{bmatrix} \begin{array}{l} \geqslant 0 \\ \geqslant 0 \\ \to \min \end{array}
$$

165

第6章 运输问题

16. 第1阶段

在过去的岁月里，你是否有机会阅读过杂志呢？这些杂志里的广告是很吸引人的．在一本旧杂志里有这样一则广告：蓝色牛仔裤，农业工人最基本的服装，每条 2.99 美元．然而，广告里还有一行小字写着：

在洛矶山脉的西边，价格也许会高些．

为什么这些位于东海岸(East Coast)的蓝色牛仔裤生产商要把不承担责任的声明写在他们的广告里？显然，他们面临着使人畏缩的任务和相应的费用，那就是将产品经过洛矶山运到洛矶山以西需要蓝色牛仔裤的消费者手中．现在随着生产厂家遍布全美而不是只在东部，以及运输手段的提高，在广告里再加上这类不承担责任的声明可能是没有必要的．但是，对工商业而言，运输费用仍是考虑的重要因素．

第1章的例 2.4 讨论了运输问题的一个特例并用反复试验的方法解决了此问题．因为运输问题为线性规划提供了特别简单和好的例子，我们将再次回顾这些例子，把它们化成线性规划模型并用刚学过的方法求解．从历史上来看，这些问题在一般的线性规划被研究之前就被详细地研究过了．有趣的是，我们在第5章考虑的对偶概念最初是出现在运输问题中的．

本章我们将看到单纯形法的第1阶段(找可行解)可以手工完成．而且，单纯形法的第2阶段(找最优解)在运输问题中比一般的线性规划简单．

为了引入图形和表格，我们考虑例 15.7 并从图形表示开始(见图 16.1)．

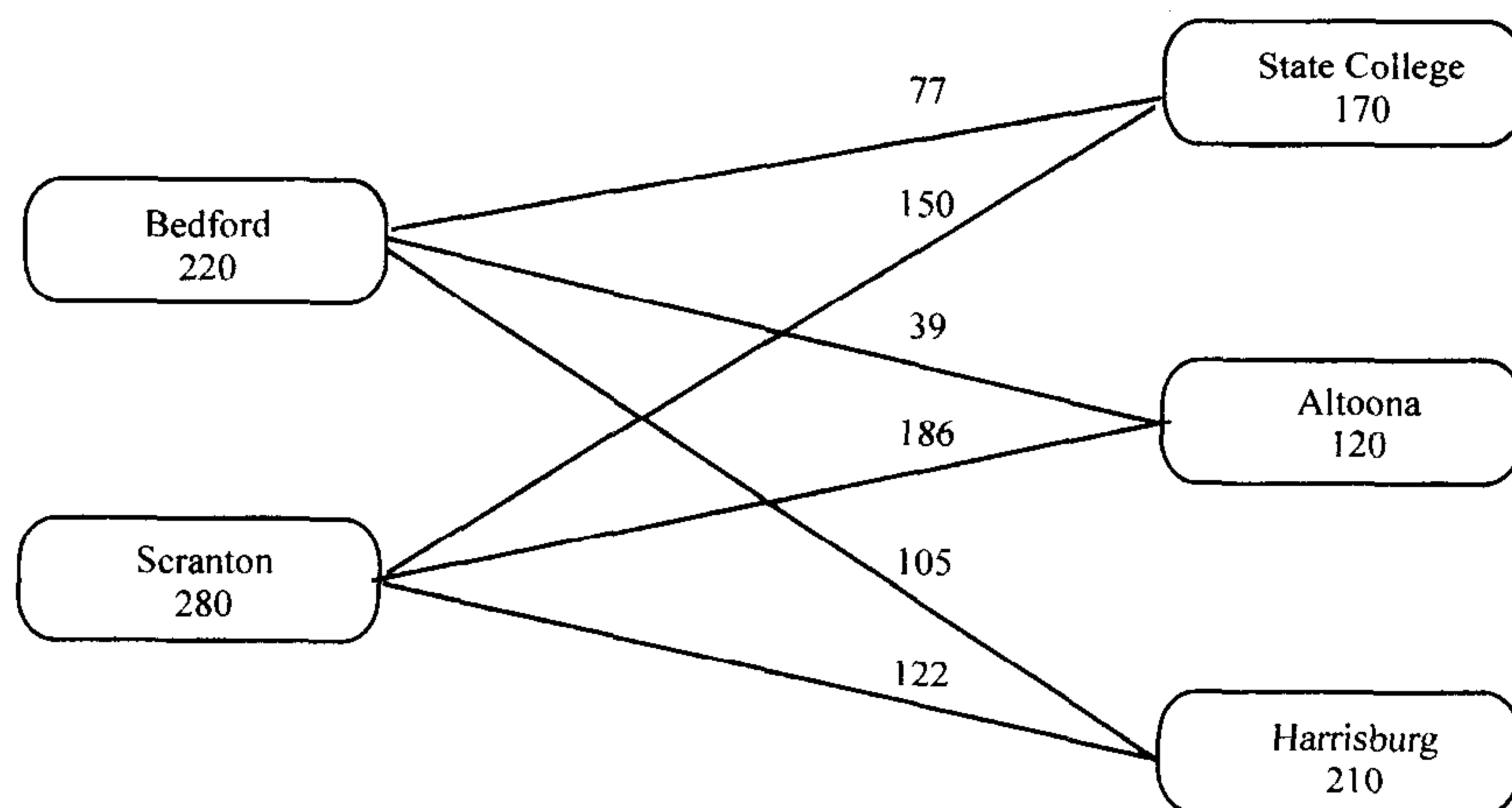


图 16.1　例 15.7 的图示

相关数据见表 16.2.

表 16.2 例 15.7 的数据表格

77	39	105	220
150	186	122	280
170	120	210	

我们在此表中为未知数 x_{ij} 留有空格，它们有时被叫做流(flow). 任何运输问题都可类似地处理. 注意运输问题不存在可行解，除非以下*平衡条件*成立：总的供应(右端数之和)与总的需求(底部的数之和)相等. 例 2.4 是一个开放的运输问题，即平衡条件不成立. 但这个问题可以写成一个满足平衡条件的封闭的运输问题，只需引入一个虚构的商店，*假需求点*(dummy demand point)，过剩的产品将免费地运到这个点. 如果我们假设虚构商店的需求量恰好是多余的供应量，那么平衡条件就成立.

167

表中位置对应着图示中的箭头，在表中的(i, j)位置填入一个数对应着从仓库 i 向商店 j 供货. 下面是一种寻找可行解的方法. 首先验证平衡条件，它对这个问题是满足的，所以可以继续. 在表中选取一个位置并写上一个最大可能的数，即相应仓库的可用货物.

例如，在例 15.7 中我们选取第一行的第一个位置并写上 170.

170			~~220~~ 50
			280
~~170~~	120	210	

我们不写出目标函数，因为还不用它. 注意表的第一列已完成，它对应着 State College 的商店. 通过删去需求 170 来表示. 因为这个需求已经满足，第一列的第 2 个元素现在必须是 0，不用将 0 写出来. 我们在表的右端填上一个新数 50 代替 200 以追踪没有分配的仓库存货.

于是我们填上一个数，删去一个数并调整另一个数. 第一列已经做完(我们在里面找到一个流)，我们将在一个 2×2 的小表上继续做下去.

在表中选取第二个位置并写上一个最大可能的数 50＝min(50，120)：

170	50		~~220~~ ~~50~~
			280
~~170~~	~~120~~ 70	210	

168

现在第一行和第一列都已经做完．只要还有选择余地，我们就对余下的 1×2 矩阵重复上面的过程：

170	50		~~220~~ ~~50~~
	70		~~280~~ 210
~~170~~ 0	~~120~~ ~~70~~	210	

最后一步分配将由前面的选择确定：

170	50		~~220~~ ~~50~~
	70	210	~~280~~ ~~210~~
~~170~~	~~120~~ ~~70~~	~~210~~	

对应的总费用是

$$77.170 + 39.50 + 186.70 + 122.210.$$

下面是得到的可行解，有两个 0 没写上：

170	50		220
	70	210	280
170	120	210	

这种寻找可行解的方法对任何满足平衡条件的运输问题都是可行的．下面我们叙述任意的运输问题．

一般的运输问题

一个生产商有 m 个仓库和 n 个零售商店．仓库 i 有 a_i 个单位的可用产品且商店 j 需要 b_j 个单位的产品．

169

假设下面的平衡条件成立：

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$$

例如，在例 15.7 中此平衡条件满足．正如我们所看到的，例 2.4 不满足．由仓库 i 向商店 j 运输一个单位货物的费用记作 c_{ij}，所运的数量记作 x_{ij}．此线性规划可以写成如下的形式：

$$
\begin{cases}
\text{minimize} & C(x_{11},\cdots,x_{mn}) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \\
\text{subject to} & \sum_{i=1}^{n} x_{ij} = b_j, \quad j = 1,\cdots,m \\
& \sum_{j=1}^{m} x_{ij} \leqslant a_i, \quad i = 1,\cdots,n \\
& x_{ij} \geqslant 0, i = 1,\cdots,n, j = 1,\cdots,m,
\end{cases}
$$

其中 C 是需最小化的总费用函数，约束

$$
\sum_{i=1}^{n} x_{ij} = b_j, \quad j = 1,\cdots,m
$$

表示商店 j 从仓库 i 接收的货物数量，其余的约束表示仓库 i 能提供给商店 j 的数量和通常的非负约束．

本章我们将说明怎样用单纯形法求解这个问题．现在，有一些问题需要考虑：

(a)我们已经看到最小化运输费用的方法．然而，最大化运输费用而不是最小化它们有意义吗？考虑这样一种情形：你在一家洲际运输公司工作并因此得到报酬，此时你希望运输费用最大还是最小呢？

170

(b)这个一般的运输问题足够一般吗？例 2.4 是这个一般的运输问题的特例吗？换言之，例 2.4 中的约束能写成等式而不是不等式吗？注意在此例中总供应量是 130，大于需求量 100. 因此，如果我们只用等式来代替不等式，那么我们将没有任何可行解．

注 16.3　历史上，运输问题是第一个从形式上被明确地描述，从理论上被研究、求解并在工业中使用的线性规划．

现在我们说明上面寻找可行解的方法适用于一般的运输问题．任选一个位置并写上最大可能的元素——即右端和底端数的最小者．然后删去与已写出元素相等的那个数，并调整另一个数．在最后一步，只留下 1×1 表时，我们删去这两个数．于是总步数是 $m+n-1$，其中 m 是行数，n 是列数．我们的方法给出一个可行解，它在 mn 个变量中最多有 $m+n-1$ 个非 0 元．

为了继续寻找最优解，将所选择的位置数准确地定为 $m+n-1$ 是方便的．其方法是在两个数相同的情况下(除非这是最后一步)，仅删去其中一个边际元素(并令另一个为 0)；若边际元素有一个是 0，则仍然将它写进选定的位置．通过这种方法，除最后一步外每步删去一行或一列．我们不应该删去最后残余的行或列，除非它是最后一步，即第 $m+n-1$ 步．

我们用此法得到的 $m+n-1$ 个位置对应于可行的标准行表中的基变量(在右端的变量)．在表中我们写上基变量的值，不写非基变量对应的 0，以便为对偶变量留下位置．

这样，对那些可用手工方式写下数据的小规模问题，可以很容易地找到一个可行解．事实上，我们所做的正是单纯形法的第1阶段．我们使用较小的表格而不是标准表．开始时，所有的变量 x_{ij} 都在顶端，松弛变量在右端．选中的位置 (i, j) 对应着我们要旋转到右端的变量 x_{ij}，旋转元是1或−1.

171

例 16.4 为下面的运输问题找一个基可行解：

			2
			3
2	2	1	

因为找一个可行解不需要目标函数，所以没有给出目标函数．我们用西北角(northwest)方法．根据此法，选择左上角的位置．注意我们删去第1行的2并调整右端的2为0：

2			~~2~~ 0
			3
~~2~~	2	1	

现在第1列已做完，西北角方法告诉我们选择第1行的第2个位置并填上可能的最大数0：

2	0		~~2~~ ~~0~~
			3
~~2~~	2	1	

现在第1行和第1列都已做完．我们要选择的下一个位置在第2行和第2列：

2	0		~~2~~ ~~0~~
	2		~~3~~ 1
~~2~~	~~2~~	1	

现在只剩下最后一行和列．我们写上最后元素1并删去边上的最后两个数：

2	0		~~2~~ ~~0~~
	2	1	~~3~~ ~~1~~
~~2~~	~~2~~	~~1~~	

此例中，$m=2$，$n=3$，我们在表中写上 $m+n-1=4$ 个数，其中有一个是0. ■

用图论的术语来讲，所选的位置是连接所有顶点(节点)的 $m+n-1$ 条边(弧)，形成一个称为树的图(网络)．这棵树正好有 $m+n-1$ 个箭头(或边)，是连通的且没有环(圈).

172

像前面提到的一样，如果这个问题被表示成一个标准行表，所选的位置对应着行可行表右端的基变量，其他位置对应着行可行表顶端的非基变量．这种方法事实上实施的是单纯形法的第 1 阶段．没有必要对运输问题的整个表进行操作，因为表的 a 区总是由 0，1，－1 组成，表的 b 区由表中所选位置的值组成，c 区的构造将在下面讨论．

如果平衡条件成立，我们总可以得到一个可行解，否则我们知道问题不可行，停止．这是我们在选取元素之前检验平衡条件的原因．若供应和需求数据都是整数，用此方法得到的可行解也是整数．它总是可行域的顶点．证明留作练习．

其实我们有许多选择表中位置的方法．西北角方法忽略了目标函数，其他根据目标函数选择位置的方法可以得到一个更好的可行解．例如，在最小费用方法中，我们寻找一个有费用最小的位置．

为了进行单纯形法的第 2 阶段，我们必须从表的 b 区即从选出构造可行解的表的元素上重新恢复表的 c 区．所用的方法可以利用势的概念得到最好的描述．我们已经在例 15.7 中引入了势．

在一般的运输问题中，对偶问题写成

$$\text{maximize} -\sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j$$

$$\text{subject to}$$

$$w_{ij} = c_{ij} + u_i - v_j \geqslant 0 \; \forall \, i,j$$

$$u \geqslant 0, v \geqslant 0.$$

在例 15.7 中，

173

$$[a_1,a_2]=[220,280],[b_1,b_2,b_3]=[170,120,210],[c_{i,j}]=\begin{bmatrix}1 & 2 & 3\\1 & 2 & 2\end{bmatrix}.$$

势对应着每个零售商店和仓库的约束(或相应的松弛变量)，而对偶问题中的其他变量 w_{ij} 对应着原问题的决策变量 x_{ij}．它们是根据势写出的对偶问题的松弛变量．我们称这些变量 w_{ij} 为*偏差*(discrepancy)．

开始时，所有的 x_{ij} 在标准表(见例 15.7)的右端，w_{ij} 在左端．经过几次旋转后，对应于被选中的位置(i，j)的变量 x_{ij} 转到了可行表的顶端，并且相应的 w_{ij} 转到了底端．它们当前的值写在表的最后一行，表的“c 区”.

为了进行单纯形法的第 2 阶段，我们必须从表的 b 区上重新恢复表的 c 区．我们知道，如果 x_{ij} 是行问题的基变量，即它出现在表的右端，那么相应的对偶变量 w_{ij} 出现在左端且 w_{ij} 在列问题的基解中取 0(包括势)．

于是，对所有选中的位置(i，j)有 $w_{ij}=c_{ij}+u_i-v_j=0$. 可以证明这些方程唯一地确定势，除了相差一个常数之外．这点可以从如下事实得到证实：选中的位置确定了图形表示(图 16.5)中的一棵树．

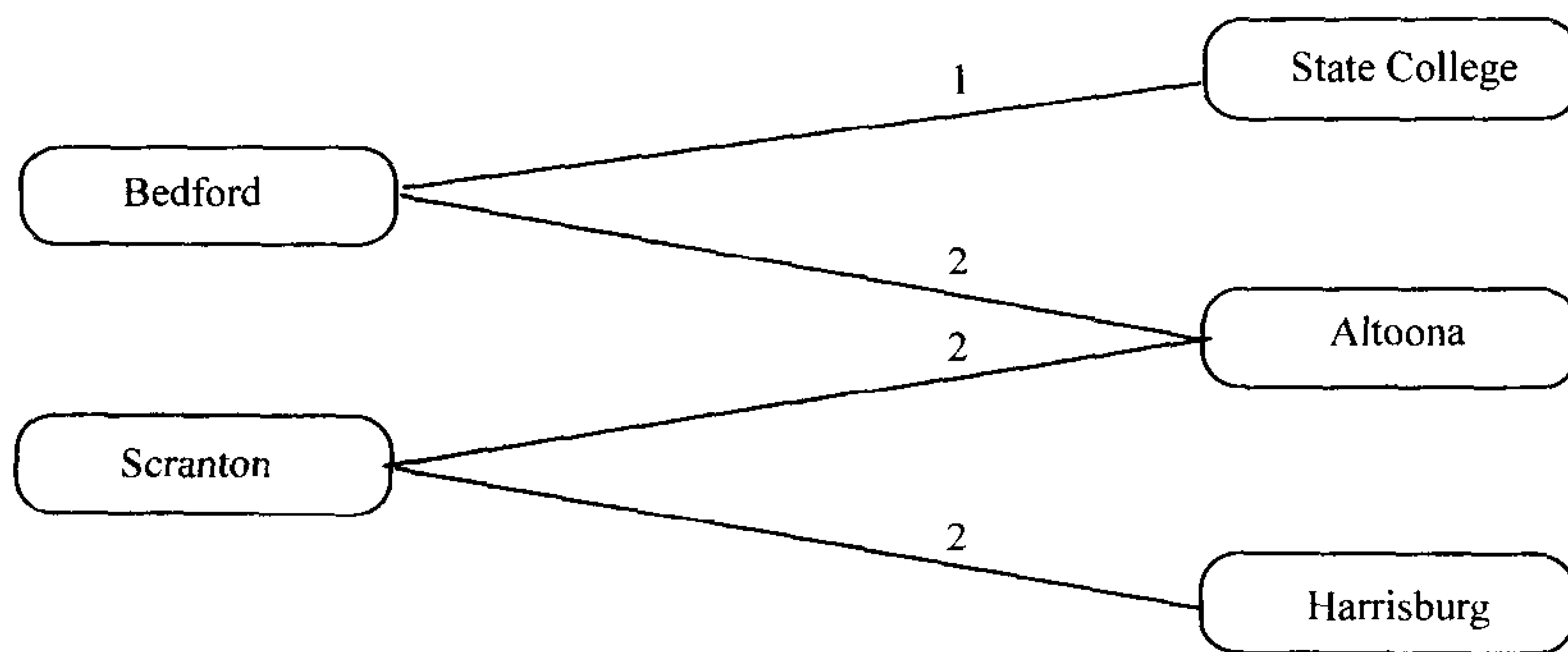


图 16.5　例 15.7 可行解表示的树

[174]

继续考察例 15.7. 我们在被选取用做可行解的位置上标上一个 * 号(表 16.6).

表 16.6　例 15.7 的可行解的基础位置

77 *	39 *	105	220
150	186 *	122 *	280
170	120	210	

树的边对应着选中的位置且每单位的费用写在上面．因为我们寻求的势与解可能相差一个常数，所以可以在开始时设定 $u_1=0$. 此时由方程 $c_{11}=v_1-u_1$ 得 $v_1=77$. 类似地，由方程 $c_{12}=v_2-u_1$ 得 $v_2=39$. 在 c_{22}的方程中利用 v_2 的值得到 $u_2=-147$，代入 c_{23}的方程中得到 $v_3=-25$. 于是基于 $u_1=0$ 的选择我们可以找到所有的势．显然，方程组所有的其他解可以通过给所有的势加上一个同样的常数得到．我们在图 16.7 和表 16.8 中标出这些势．

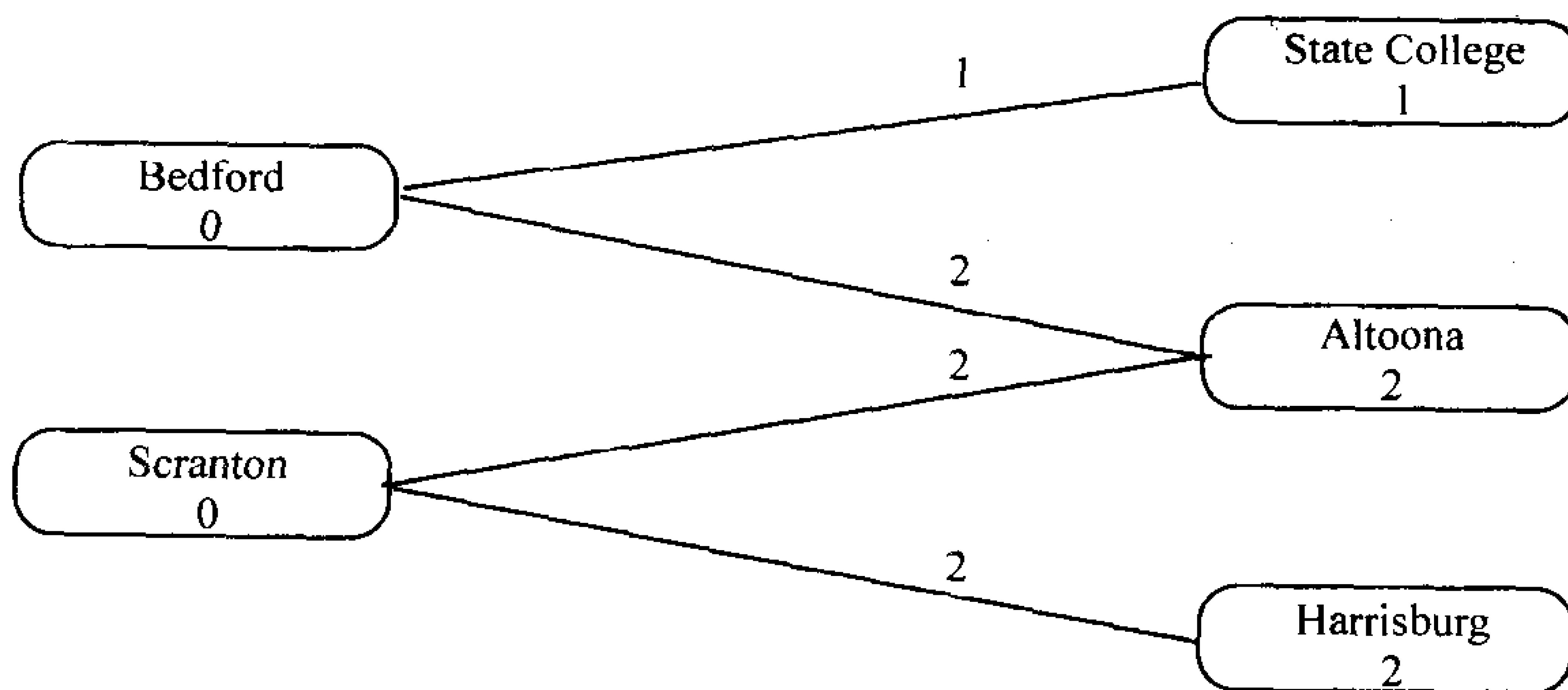


图 16.7　例 15.7 的势的树表示

[175]

表 16.8 例 15.7 的可行解的势

	77	39	−25	
0	77 *	39 *	105	220
−147	150	186 *	122 *	280
	170	120	210	

现在我们可以计算其他(非基)位置的值 $w_{ij}=c_{ij}+u_i-v_j$. 将它们写进表 16.8 并用圆括号括起来(表 16.9). 因为表中有一个 w_{ij} 是负的，它不是最优的. 我们可表示出原问题的基解，即选择位置的时候，用赋给 x_{ij} 的值代替用 * 标出的位置.

表 16.9 例 15.7 的对偶问题的基解

	77	39	−25	
0	77 *	39 *	105 (130)	220
−147	150 (−74)	186 *	122 *	280
	170	120	210	

如同前面所言，除代表目标函数当前值的最后一个元素外，w_{ij} 的值对应着表的最后一行的元素，虽然这一行只记在我们的脑海中而没有写出来. 现在我们已经为旋转做好了准备.

练习

1～3. 为下列每个运输问题找一个可行解.

1.

				35
				20
20	10	10	15	

2.

				35
				9
				11
				20
40	10	10	15	

3.

				35
				90
				111
				20
140	91	10	19	

4～6. 求可行解和相应的总运输费用.

4.

1	2	3	1	2	3	1	2	3	10
0	3	2	0	1	2	1	2	1	20
2	1	0	1	2	1	2	1	1	12
1	1	1	2	2	2	2	2	1	8
2	3	4	5	1	2	3	14	16	

177

5.

1	2	3	1	2	3	1	2	3	15
0	3	2	0	1	2	1	2	1	15
2	1	0	1	2	1	2	1	1	12
1	1	1	2	2	2	2	2	1	8
2	8	4	15	11	2	3	4	1	

6.

1	2	3	1	2	3	1	2	3	15
0	3	2	0	1	2	1	2	1	5
2	1	0	1	2	1	2	1	1	12
1	1	1	2	2	2	2	2	1	18
12	8	14	5	1	2	3	4	1	

7. 例 2.4.

55	30	40	50	40	0	50
35	30	100	45	60	0	30
40	60	95	35	30	0	50
25	10	20	30	15	30	

[178] 8. 证明由我们的方法得到的可行解是可行域的一个顶点.

17. 第 2 阶段

如果计算势后，某个 w_{ij} 是负的，则根据单纯形法的第 2 阶段，我们应从基中剔除 w_{ij} 并引入另一个变量代替它. 对原问题来说，x_{ij} 必须进基代替另一个变量. 此时，目标函数值应该得到改进或至少保持不变. 从图形上看，基的改变对应着在树上加一条新边，并移走因此出现圈中的一条边形成一棵新的树. 我们将用例 2.4 说明这一过程. 首先用西北角方法求一个基可行解(表 17.1).

表 17.1 例 2.4 的第一个可行解

55 25	30 10	40 15	50	40	0	50
35	30	100 5	45 25	60	0	30
40	60	95	35 5	30 15	0 30	50
25	10	20	30	15	30	

相应的运输费用是 25×55+10×30+15×40+5×100+25×45+5×35+15×30+30×0=1 375+300+600+500+1 125+175+450+0=4 525. 我们将看

到，忽略目标函数会导致需要许多次旋转才可得到最优表．通常，在解运输问题的第 1 阶段，多花费一些时间是值得的，因为我们在第 2 阶段可以少花费时间．同前面一样，我们寻找势(表 17.2).

表 17.2　表 17.2 的势

	55	30	40	−15	−20	−50	
0	55 25	30 10	40 15	50	40	0	50
−60	35	30	100 5	45 25	60	0	30
−50	40	60	95	35 5	30 15	0 30	50
	25	10	20	30	15	30	

179

我们随机地对第一行取第一个势为 0. 运输费用也可以由势而不是由流计算得到：55×25+30×10+40×20−15×30−20×15−50×30−(0×50−60×30−50×50)=4 525. 和前面的例子一样，可行解确定了图论表示中的一棵树(图 17.3).

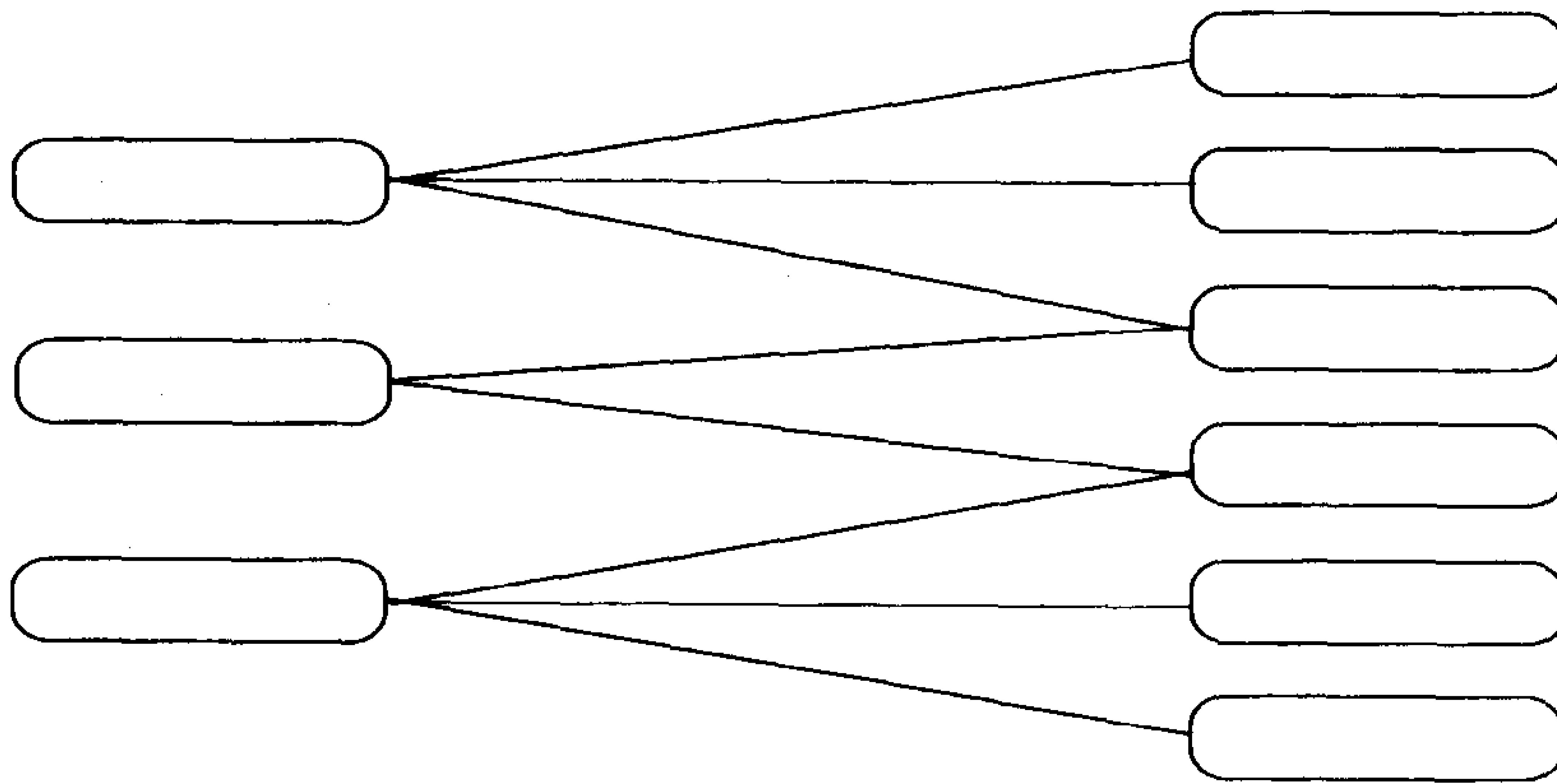


图 17.3　表 17.1 的树

下面我们计算对偶松弛变量(偏差)w_{ij}的值(表 17.4). 在表中每个非基变量 w_{ij} 出现在左端，而对应的基变量 x_{ij} 出现在右端，w_{ij} 在这个基解中的值是 0. 基变量 w_{ij} 对应着非基变量 x_{ij} 且出现在表的底端，故其值属于最后一行．偏差 w_{ij} 由表 17.1 中选取的位置确定．前面介绍过，选中位置上的偏差是 0 且不写出来，而其他的偏差写在圆括号里．换言之，偏差和流一样，是由这棵树确定的(见图 17.3).

表 17.4　表 17.2 的偏差 w_{ij}

	55	30	40	−15	−20	−50	
0	55 25	30 10	40 15	50 (65)	40 (60)	0 (50)	50
−60	35 (−80)	30 (−60)	100 5	45 25	60 (20)	0 (−10)	30
−50	40 (−65)	60 (−20)	95 (5)	35 5	30 15	0 30	50
	25	10	20	30	15	30	

[180]

有些基变量 w_{ij} 是负的，所以解不是最优的．我们必须选取一个负的基变量 w_{ij} 和一个非基变量 $w_{i'j'}$ 交换，这通过交换非基变量 x_{ij} 和基变量 $x_{i'j'}$ 在表上实现．我们选取 $w_{21}=-80$，并决定把 x_{21} 和一个基变量交换．用图论的术语讲，我们增加一条从仓库 2 到商店 1 的边，这时产生了圈(2，1)，(2，3)，(1，3)，(1，1)，从而破坏了图的树结构(图 17.5)．

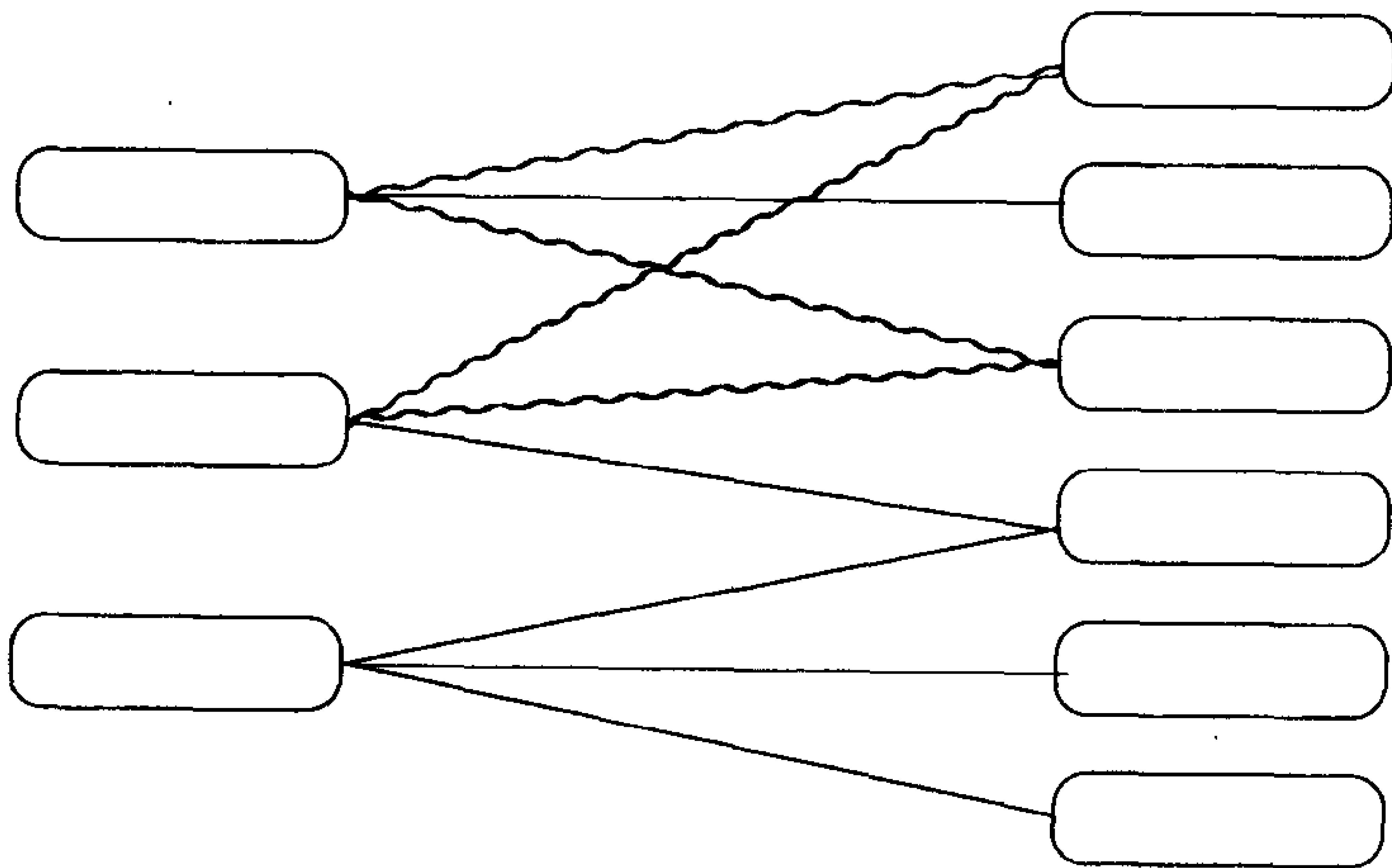


图 17.5　x_{21} 进基产生循环

在由新选取的仓库 2 到商店 1 的边上，从一个 $\varepsilon\geqslant 0$ 的流开始，即令 $x_{21}=\varepsilon$. 我们必须调整圈上的流，以便满足平衡约束．因此从 x_{23} 减去 ε，使其不超过仓库 2 的供应量．然后对 x_{13} 加上 ε，以便满足商店 3 的需求．在圈的最后一条边上，将 x_{11} 减去 ε，以免超出仓库 1 的供应量和商店 1 的需求量(表 17.6)．

表 17.6　通过 ε 调整圈上的流

	55	30	40	−15	−20	−50	
0	55 25−ε	30 10	40 15+ε	50 (65)	40 (60)	0 (50)	50
−60	35 (−80)ε	30 (−60)	100 5−ε	45 25	60 (20)	0 (−10)	30
−50	40 (−65)	60 (−20)	95 (5)	35 5	30 15	0 30	50
	25	10	20	30	15	30	

181

运输费用的变化量是$(35-100+40-55)\varepsilon=-80\varepsilon$. 注意系数$-80$正是$w_{ij}$的值！为了使目标函数得到最好的改进，我们让$\varepsilon$取最大可能的值——5. 这就使得$x_{ij}$不再被选定，或等价地从图上移去对应的边，重新恢复成一棵树．令x_{21}进基，x_{23}出基，以便得到一个新的基解(表 17.7).

表 17.7　例 2.4 的第二个可行解

55 20	30 10	40 20	50	40	0	50
35 5	30	100	45 25	60	0	30
40	60	95	35 5	30 15	0 30	50
25	10	20	30	15	30	

现在运输费用是$20\times55+10\times30+20\times40+5\times35+25\times45+5\times35+15\times30+30\times0=4\,125=4\,525-80\times5$. 继续下去，为对偶变量找到一个新的值(表 17.8).

表 17.8　表 17.8 的势和偏差 w_{ij}

	55	30	40	65	60	30	
0	55 20	30 10	40 20	50 (−15)	40 (−20)	0 (−30)	50
20	35 5	30 (20)	100 (80)	45 25	60 (20)	0 (−10)	30
30	40 (15)	60 (60)	95 (85)	35 5	30 15	0 30	50
	25	10	20	30	15	30	

182

我们用x_{21}代替x_{23}作为基变量，进行一次旋转，得到一棵新树(图 17.9).

一些w_{ij}仍然是负的，故表不是最优的．其中有 3 个负元可供选择．选取$w_{14}=-15$，相应地在图上增加一条从仓库 1 到商店 4 的边，产生了圈(1, 4)，

(1，1)，(2，1)，(2，4).

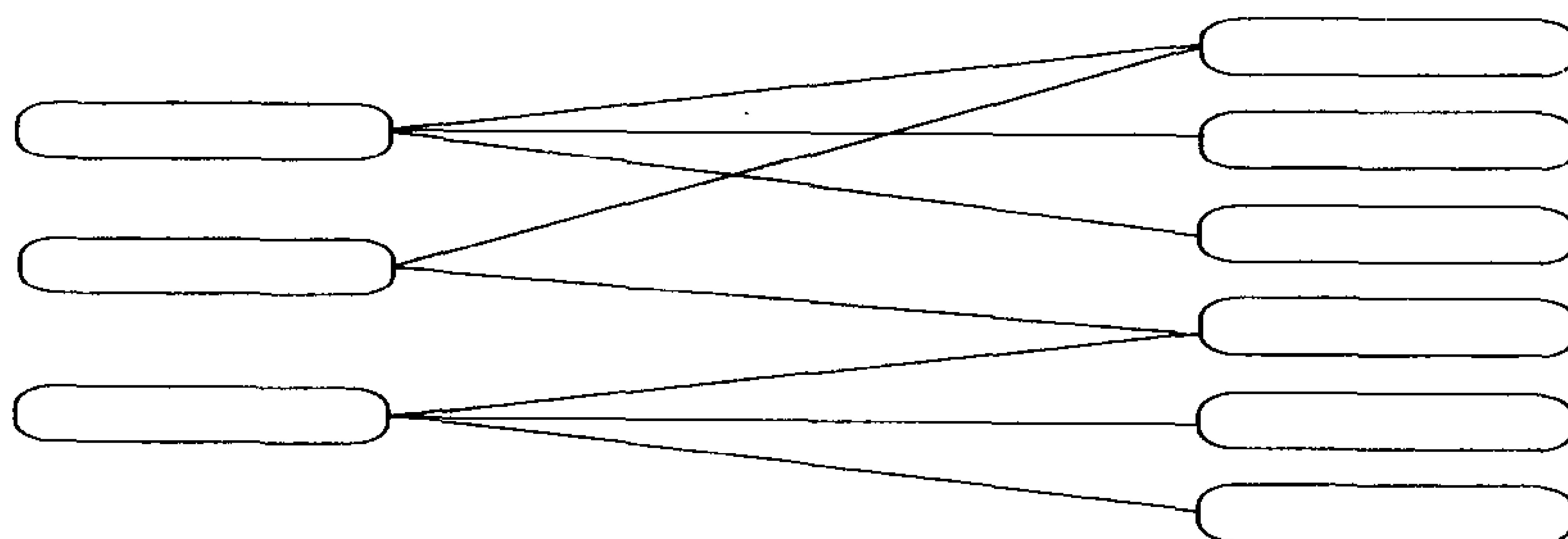


图 17.9 x_{23}出基的结果

选取 $w_{15}=-20$，相应地在图上增加一条从仓库 1 到商店 5 的边，产生了圈 (1，5)，(1，1)，(2，1)，(2，4)，(3，4)，(3，5)（图 17.10）.

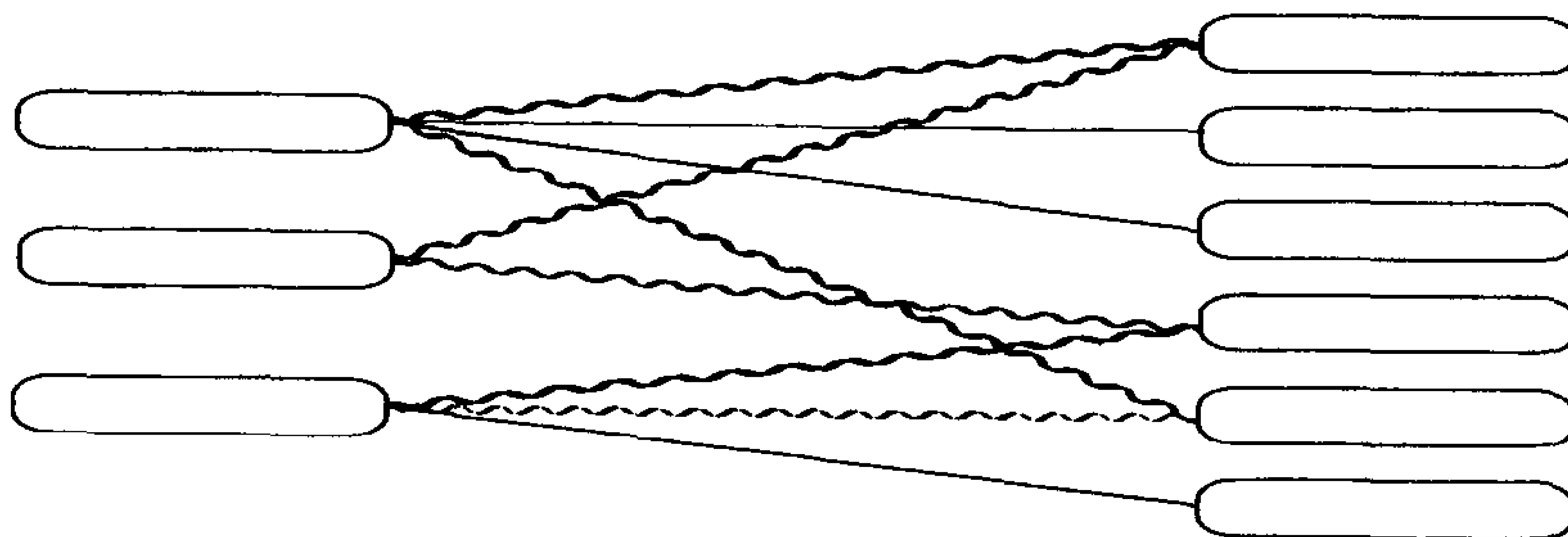


图 17.10 x_{15}进基所产生的圈

为了说明用一个较长的圈寻找一个新的基可行解的过程，我们沿后者继续下去．在由仓库 1 到商店 5 的边上从一个值为 ε 的流开始，即令 $x_{ij}=\varepsilon$. 然后令 x_{11}减去 ε，因为不能超过仓库 1 的供应量．再令 x_{21}加上 ε，以便满足商店 1 的需求量．

183

我们沿着此圈继续，令 x_{24}减去 ε，x_{34}加上 ε，最后令 x_{35}减去 ε（见表 17.11）.

表 17.11 按 ε 调整新的圈上的流

	55	30	40	65	60	30	
0	55 20−ε	30 10	40 20	50 (−15)	40 (−20) ε	0 (−30)	50
20	35 5+ε	30 (20)	100 (80)	45 25−ε	60 (20)	0 (−10)	30
30	40 (15)	60 (60)	95 (85)	35 5+ε	30 15−ε	0 30	50
	25	10	20	30	15	30	

目标函数值的变化是

$$\varepsilon(c_{15}-c_{35}+c_{34}-c_{23}+c_{21}-c_{11})$$

$$=\varepsilon(40-30+35-45+35-55)=-20\varepsilon=w_{15}\varepsilon.$$

此时，ε 能取的最大值是 15，这是因为有商店 5 的需求限制．由 ε=15 的流调整 x_{ij}，得到一个新的可行解，它使目标函数值下降为 4 125－20×15＝3 825（表 17.12）．在新解中 x_{15} 代替了 x_{35}，将我们的图形表示重新恢复为一棵树（图 17.13）．

表 17.12　例 2.4 的第三个可行解

55 5	30 10	40 20	50	40 15	0	50
35 20	30	100	45 10	60	0	30
40	60	95	35 20	30	0 30	50
25	10	20	30	15	30	

184

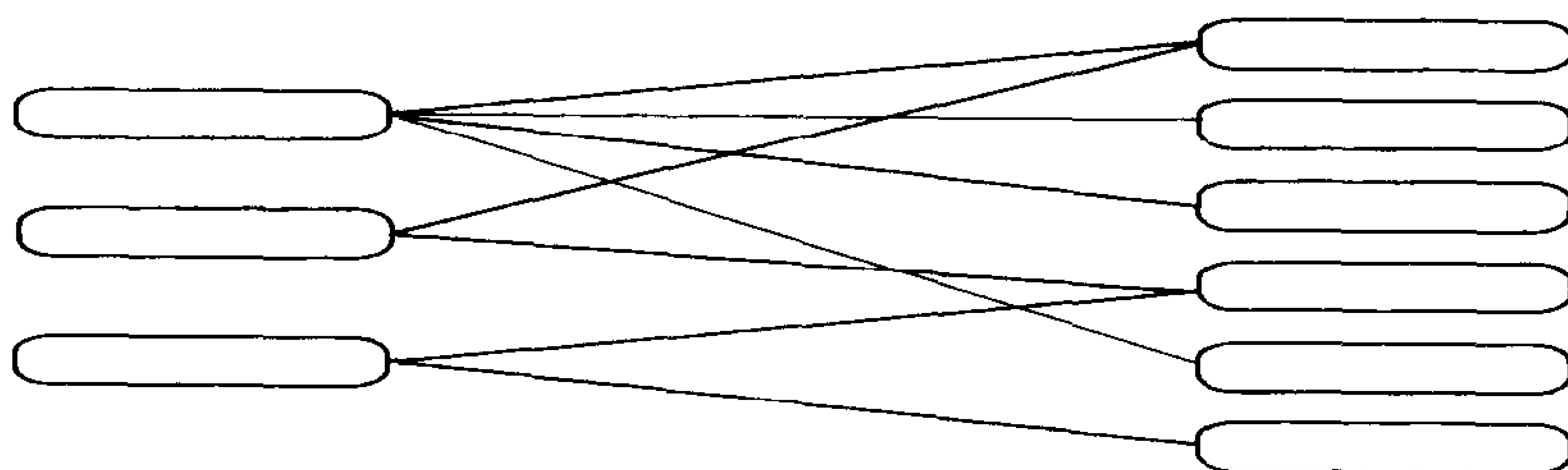


图 17.13　x_{35} 出基生成的树

同前面一样，我们继续寻找新势和新的 w_{ij}．选取负元 $w_{14}=-15$，产生圈（1，4），（1，1），（2，1），（2，4）和圈上的调整流（表 17.14）．

表 17.14　表 17.13 的对偶变量和调整流

	55	30	40	65	40	30	
0	55 5－ε ■	30 10	40 20	50 （－15） ε ■	40 15	0 （－30）	50
20	35 20＋ε ■	30 （20）	100 （80）	45 10－ε ■	60 （40）	0 （－10）	30
30	40 （15）	60 （60）	95 （85）	35 20	30 （20）	0 30	50
	25	10	20	30	15	30	

令 $\varepsilon=5$，这是受仓库 1 供应量约束的最大数，得到一个使目标函数值减少 $w_{14}\varepsilon=75$ 的新可行解(表 17.15).

表 17.15 例 2.4 的第四个可行解

55	30 10	40 20	50 5	40 15	0	50
35 25	30	100	45 5	60	0	30
40	60	95	35 20	30	0 30	50
25	10	20	30	15	30	

185

计算势和新的 w_{ij}. 取一个新的负元 $w_{16}=-15$ 和圈(1, 6), (1, 4), (3, 4),(3, 6)上值为 ε 的新流(表 17.16).

表 17.16 表 17.15 的对偶变量和调整流

	40	30	40	50	40	15	
0	55 (15)	30 10	40 20	50 $5-\varepsilon$	40 15	0 (−15) ε	50
5	35 25	30 (5)	100 (65)	45 5	60 (25)	0 (−10)	30
15	40 (15)	60 (45)	95 (70)	35 $20+\varepsilon$	30 (5)	0 $30-\varepsilon$	50
	25	10	20	30	15	30	

ε 又一次被仓库 1 的供应量限制为 5，令 $\varepsilon=5$，可以得到一个使目标函数值改进 75 的新可行解(表 17.17).

表 17.17 例 2.4 的第五个可行解

55	30 10	40 20	50	40 15	0 5	50
35 25	30	100	45 5	60	0	30
40	60	95	35 25	30	0 25	50
25	10	20	30	15	30	

186

继续计算新势和新 w_{ij}，找到一个负元 $w_{35}=-10$，产生圈(3, 5)，(3, 6)，(1, 6)，(1, 5)，令 $x_{35}=\varepsilon$，$x_{36}=25-\varepsilon$，$x_{16}=5+\varepsilon$，$x_{15}=15-\varepsilon$(表 17.18).

表 17.18　表 17.17 的对偶变量和调整流

	25	30	40	35	40	0	
0	55 (30)	30 10	40 20	50 (15)	40 15−ε	0 5+ε	50
−10	35 25	30 (−10)	100 (50)	45 5	60 (10)	0 (−10)	30
0	40 (15)	60 (30)	95 (55)	35 25	30 (−10) ε	0 25−ε	50
	25	10	20	30	15	30	

令 ε 为商店 5 的需求量限制的最大值 15，可以产生一个使运输费用减少 150 的新可行解(表 17.19).

表 17.19　例 2.4 的第六个可行解

55	30 10	40 20	50	40	0 20	50
35 25	30	100	45 5	60	0	30
40	60	95	35 25	30 15	0 10	50
25	10	20	30	15	30	

然后我们计算新势 u_i，v_j 和新的对偶松弛变量 w_{ij}. 这次我们选择位置(2，6)，$w_{26}=-10$. 令 $x_{26}=\varepsilon$，$x_{24}=5-\varepsilon$，$x_{34}=25+\varepsilon$，$x_{36}=10-\varepsilon$ 调整圈上的流，得到新表(表 17.20).

187

表 17.20　表 17.19 的对偶变量和调整流

	25	30	40	35	30	0	
0	55 (30)	30 10	40 20	50 (15)	40 (10)	0 20	50
−10	35 25	30 (−10)	100 (50)	45 5−ε	60 (20)	0 (−10) ε	30
0	40 (15)	60 (30)	95 (55)	35 25+ε	30 15	0 10−ε	50
	25	10	20	30	15	30	

令 ε 为商店 2 的供应量限制的最大值 5，可以得到一个使目标函数值改进 50 的可行解(表 17.21).

表 17.21　例 2.4 的第七个可行解

55	30 10	40 20	50	40	0 20	50
35 25	30	100	45	60	0 5	30
40	60	95	35 30	30 15	0 5	50
25	10	20	30	15	30	

下面我们计算势和对偶松弛变量即对应表的 c 区(表 17.22).

表 17.22　表 17.22 的势和 w_{ij}

	35	30	40	35	30	0	
0	55 (20)	30 10	40 20	50 (15)	40 (10)	0 20	50
0	35 25	30 (0)	100 (60)	45 (10)	60 (30)	0 5	30
0	40 (5)	60 (30)	95 (55)	35 30	30 15	0 5	50
	25	10	20	30	15	30	

最后，与一个标准表的 c 区对应的所有基变量 w_{ij} 非负，因此第七个可行解是最优的．事实上对例 2.4 这个解是用推理得到的，虽然对大型问题这种推理是很困难的．运输费用是 $10\times30+20\times40+20\times0+25\times35+5\times0+30\times35+15\times30+5\times0=3\,475$. 图 17.23 显示的是最优解的图形表示．

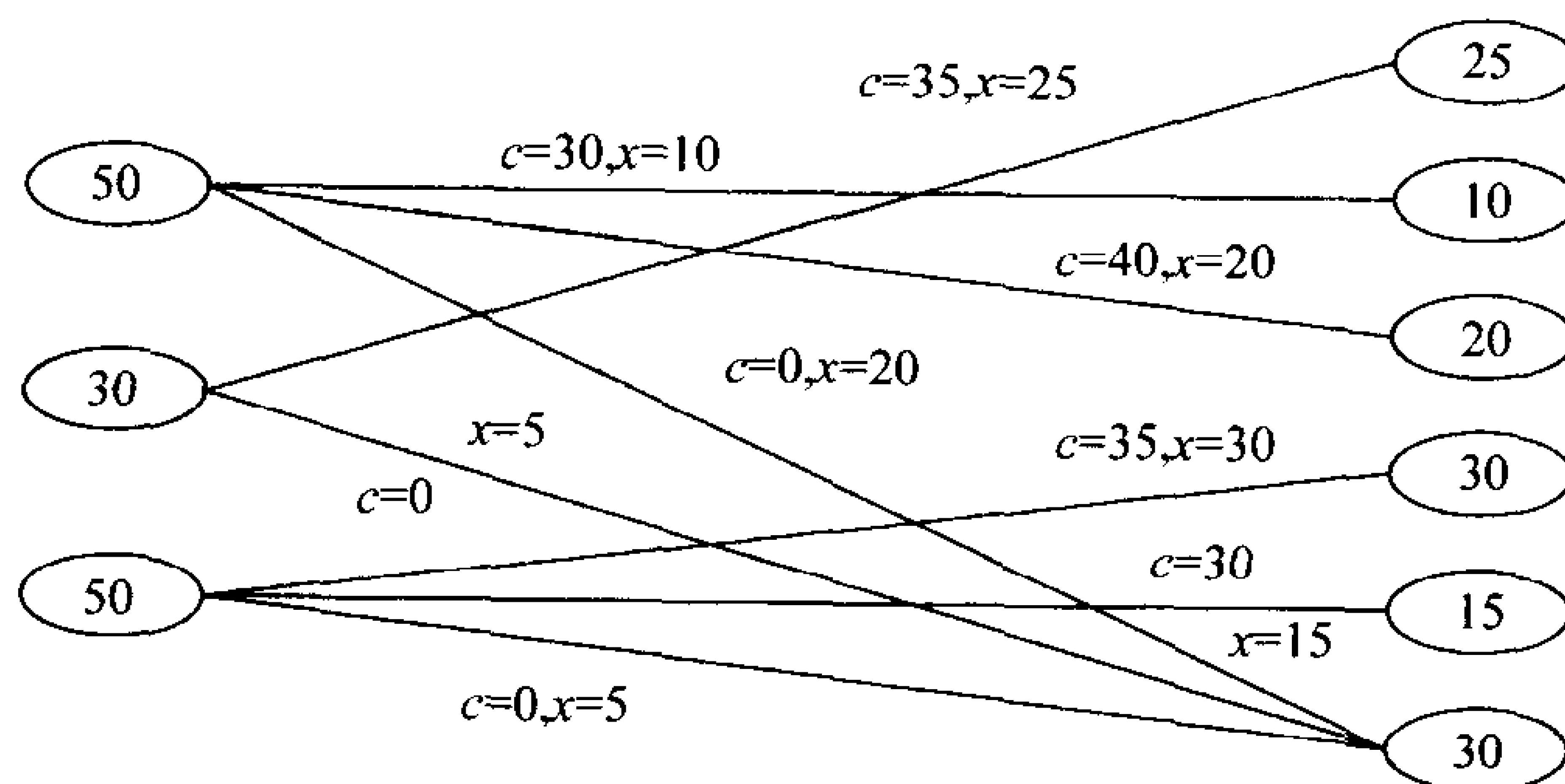


图 17.23　与表 17.22 对应的最优解

因 $w_{22}=0$，可以增加 x_{22} 而不改变目标函数．如果这个 x_{ij} 能够增加而不与任何约束冲突，就可能出现其他的最优解．我们可以在图上加一条由仓库 2 到商店 2 的边，产生圈(2，2)，(2，6)，(1，6)，(1，2)来验证此可能性．事实证明由仓库 1 到商店 2 每单位的运费等于由仓库 2 到商店 2 的费用，每单位存货的费用即从所有仓库运到商店 6 是免费的．于是我们得到结论：由仓库 2 而不是仓库 1

满足商店 2 的一些需求，让仓库 2 有少许存货而仓库 1 有更多的存货，将不会改变运输费用．我们最多能调整 5 个单位，因为仓库 2 只有 5 个单位的余量．因为其他所有的 w_{ij} 都是严格正的，所以用这种方法可以得到所有的最优解．

将一些额外的数标在图上，能表示出全部的解(图 17.24)．图的每条边由仓库 i 到商店 j 的供货数量 x_{ij} 标出．注意“商店 6”是虚拟的． 189

对每个 i，x_{i6} 实际上是仓库 i 的剩余存货的数量．我们可以标出每个仓库的总供应量和每个商店的总需求量，其中商店 6 标出的是所有仓库总的多余的供应量．通过增加一条由仓库 2 到商店 2 的边并使 x_{12} 和 x_{22} 是 ε 的函数(图 17.24)，我们可以表示出所有的最优解．

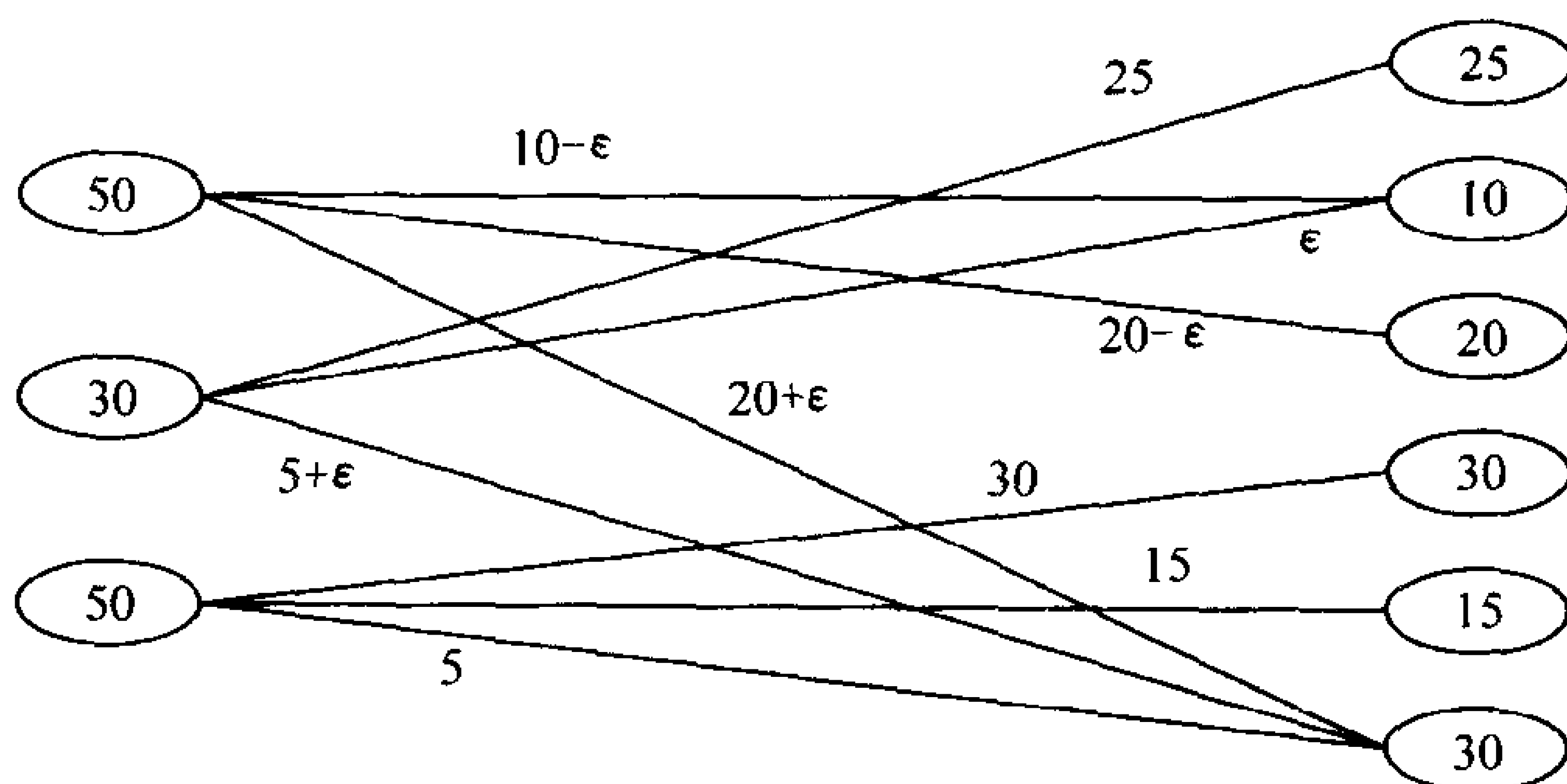


图 17.24　全部最优解，$0\leqslant\varepsilon\leqslant5$

运输问题出现循环的可能性非常小，并且 Bland 法则可以防止出现循环．若给出的供应和需求数据是整数，则最优解也是整数．而且，可行解里整数的大小不会超过供应和需求数据的整数的大小．

利用旧势计算新势是可能的．这可以减少计算量，但增加了方法的复杂性，且增加了出现新错误的机会．

注 17.25　运输问题的对偶单纯形法就是著名的 Hungarian 方法．它基于下面的思想：费用矩阵的一行或一列加上一个常数不改变最优域，因为目标函数仅改变了一个常数．如果我们设法得到一个所有元素≥0 且有足够 0 元的费用矩阵，而整个流只放在 0 费用的位置上，则我们得到一个最优解(调整后的总费用为 0)．特别建议用对偶法求解来自工作指派问题的运输问题(见下一节)，其中的退化使得原始方法很慢．在这类问题中，$m=n$ 且对所有的 i，j 有 $a_i=b_j=1$． 190

下面我们简要叙述求解这类问题的 Hungarian 方法．此方法适应于每行和每列有一个 0 元的非负矩阵．将矩阵 $[c_{ij}]$ 的每一行减去该行的最小元，我们得到一个 $n\times n$ 初始矩阵．然后对每列也做同样的运算．

1. 给定任何非负矩阵 C，对行和列画出最少数目的 t 条直线来覆盖 C 中全部 0 元．若 $t=n$，则像第 16 节那样选取有 0 费用的位置并接连地删去这 n 条线，就很容易找到一个基可行解．否则，转向第 2 步．

2. $t<n$. 计算所有没被覆盖的元素的最小值 m. 然后将每个没被覆盖的元素减去 m，每个被覆盖两次(即既被一条水平线覆盖，也被一条垂直线覆盖)的元素加上 m. 转第 1 步.

练习

1～2. 求运输问题的可行解及对应的对偶问题的基解，这里星号(＊)标出行问题的基变量. 把这两个基解放在同一表里. 它们是最优的吗?

1.

1 *	2 *	3	200
1	2 *	2 *	300
175	125	200	

2.

1	2 *	3 *	200
1 *	2 *	2	300
175	125	200	

3～5. 求解第 16 节中练习 4～6 的运输问题.

6～8. 求解例 2.4 的运输问题，要求如下:

6. 供应量 30 用 $30+t$ 代替，其中 t 是一个参数且满足 $|t|\leqslant 1$.

7. 供应量 30 用参数 t 代替.

8. 单位费用 $c_{23}=100$ 用参数 t 代替.

191 提示: 从表 17.22 开始.

18. 工作指派问题

在第 1 章中我们已经介绍了这类问题(见第 2 节例 2.5 和练习 6～9). 一般地，根据变量和约束，这类问题可叙述如下. 根据是否指派第 i 个人完成工作 j，设

$$x_{ij}=0 \text{ 或 } 1, \tag{18.1}$$

那么总时间是

$$\sum_i \sum_j c_{ij}x_{ij}\text{(极小化)}. \tag{18.2}$$

每个人 i 只指派做一个工作的约束是

$$\sum_j x_{ij}=1 \text{ 对所有 } i. \tag{18.3}$$

每个工作 j 只指派一个人去做的约束是

$$\sum_i x_{ij} = 1 \text{ 对所有 } j. \tag{18.4}$$

于是此问题就被描述成一个优化问题(18.1～18.4). 目标函数(18.2)及约束(18.3)和(18.4)是线性的，而约束(18.1)不是线性的.

考虑到(18.3)和(18.4)，约束(18.1)可由下面的约束代替：$x_{ij} \geqslant 0$，且为整数. 这是一个整数规划的例子.

如果我们去掉 x_{ij} 为整数这个条件(保留条件 $x_{ij} \geqslant 0$)，那么我们的问题从数学上看像是运输问题的一个特例，在这个运输问题中所有的供应量和需求量都是1，平衡条件意味着人数和工作数相等.

现在我们去掉整数约束，用单纯形法求解相应的线性规划. 将每个人看作一个提供工作的"仓库"而把每个工作看作一个需求工作的"商店". 设每个人的供应量和每个工作的需求量都是1，总的完工时间就成了总的运输费用. 192

因为整数数据保证整数解，我们的最优解也是整数，因此我们得到一个工作指派问题的最优解.

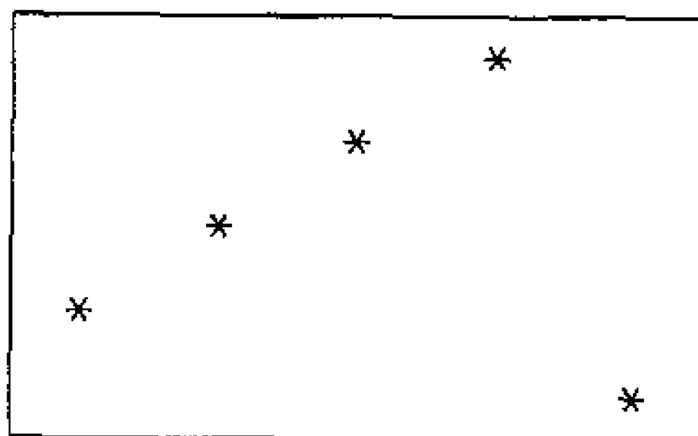
于是我们可将任意一个工作指派问题转化为运输问题. 指派问题的一个变种是最大化完成一组工作的效率和 S 而不是最小化总费用(如时间). 利用标准技巧，我们可将求最大转化为求最小. 导出的运输问题也许有非系数 c_{ij}，这在前面的例子中没有出现. 然而，无论系数是正，负，或0，或是它们的混合，单纯形法都可以求解. 因此即使有这种情况出现我们也可以像前面一样求解.

问题 18.5 求解工作指派问题

2	2	2	1	2
2	3	1	2	4
2	0	1	1	1
2	3	4	3	3
3	2	1	2	1

表中数据是每人每个工作的费用，目的是求最小总费用.

解 除非你喜欢旋转，否则在第1阶段不使用西北角方法，而是多花费些时间，这是值得的. 我们考察每人的最小费用，即每行的最小费用：1，1，0，2，1，所以最小的总费用至少是5(即 min$\geqslant$5). 我们在每行选择一个有最小费用的位置，设法达到这个费用. 除了最后一行有两个最小费用为1的位置以外，这种位置在其他的行都是唯一的. 但它们中有一个与第2行选中的位置抵触，故我们选取最后一行的最后一个位置. 于是，我们得到一个最优解且它是唯一的：



注意此题共有 5! =120 个可行解，全部试一遍需要比我们利用常识的方法花费更多的时间．■

[193]

问题 18.6 求解问题 18.5 中的工作指派问题，但现在的数据解释为每个工人在每个工作上的效率，目的是使效率之和最大．

解 使用旋转之前，我们利用常识尝试一下．首先我们为每个人选择一个最好的工作(这在前面的问题中效果很好)，这是行中最大元的模式：

*	*	*		
				*
*				
		*		
	*			

我们得到 max≤2+4+2+4+3=15，但是如果我们在每行中选择最好的数将无法完成工作 4. 然后我们设法在每列选择一个最大元：

	*			*
	*	*	*	
*				

现在第 1 个和第 3 个工人无事可做，而我们的最大值的上界是 max≤3+3+4+3+4=17，它弱于前面的界．

事实上，因为每行选择最大元是无法完成工作的，我们得到 max<15. 因为最大值是一个整数，我们可知 max≤14. 现在我们再看一下数据和每行最大元的模式，容易看到一个总效率为 14 的可行解：

	*			
				*
*				
		*		
			*	

[194]

这里工人 1～4 每人做了最好的工作，工人 5 做了第 2 个最好的工作．■

问题 18.7 最小化匹配问题中的总损失．

1	1	2	1
2	3	2	3
4	3	3	4
4	4	5	6

解 取每列的最小元

*	*	*	*
		*	

不能产生任何可行解，而我们得到界 min>5. 取每行的最小元

*	*		*
*		*	
	*	*	
*	*	*	*

产生四个最优解. 然而，作为一个变化，我们忽略最优解而使用单纯形法. 这个练习的要点是虽然常识和特殊技巧可以对一些问题给出捷径，但单纯形法适应于每个线性规划.

我们将这个问题当作一个运输问题解决. 现在取 2000 年一名学生发现的解作为最初的基解：

$^{1}1$	^{1}	$^{2}0$	^{1}	~~1~~ θ
^{2}	^{3}	^{2}	$^{3}1$	~~1~~
^{4}	$^{3}1$	^{3}	^{4}	~~1~~
^{4}	$^{4}0$	$^{5}1$	$^{6}0$	~~1~~ θ
~~1~~	~~1~~ θ	~~1~~	~~1~~ θ	

注意这个解的总费用是 1+3+3+5=12，这比西北角方法得到的 1+3+3+6=13 好.

195

然后我们找到势和 w_{ij}：

	5	5	6	7
4	$^{1}1$	$^{1}(0)$	$^{2}0$	$^{1}(-2)$
4	$^{2}(1)$	$^{3}(2)$	$^{2}(0)$	$^{3}1$
2	$^{4}(1)$	$^{3}1$	$^{3}(-1)$	$^{4}(-1)$
1	$^{4}(0)$	$^{4}0$	$^{5}1$	$^{6}0$

顶端的第 1 个势 5 是任选的，现在有三个负的 w_{ij}. 我们选取 $w_{33}=-1$ 的位置(3，3)(还有其他两个可能的选择)，这产生了一个长度为 6 的圈. ($w_{14}=-2$ 的选择将导致一个长度为 4 的圈和一个退化的旋转步). 下面是沿此圈流的一个调整：

$^{1}1$	$^{1}(0)$	$^{2}0$	$^{1}(-2)$
$^{2}(1)$	$^{3}(3)$	$^{2}(0)$	$^{3}1$
$^{4}(1)$	$^{3}1-\varepsilon$	$^{3}\varepsilon$	$^{4}(-1)$
$^{4}(0)$	$^{4}0+\varepsilon$	$^{5}1-\varepsilon$	$^{6}0$

令 $\varepsilon=1$ 并选定位置(4，3)(另一个可能的选择是位置(3，2))，目标函数改进 1. 下面是新的基可行解、新的势以及沿一个长度为 6 的圈流的新调整：

	5	6	6	8
4	$^{1}1$	$^{1}(-1)$	$^{2}0-\varepsilon$	$^{1}(-3)\varepsilon$
5	$^{2}(2)$	$^{3}(2)$	$^{2}(1)$	$^{3}1$
3	$^{4}(2)$	$^{3}0-\varepsilon$	$^{3}1+\varepsilon$	$^{4}(-1)$
2	$^{4}(1)$	$^{4}1+\varepsilon$	$^{5}(1)$	$^{6}0-\varepsilon$

选择位置(1，4)作为一个新基位置(还有两种别的选择)．我们必须取 $\varepsilon=0$，所以它是一个不改变可行解的退化旋转步(但它改变了基，即基变量的集合)．现在我们选定位置(4，4)(别的可能的选择是(1，3)和(3，2))，再次计算新势和 w_{ij}，并找到一个负的 w_{ij}(一旦我们找到一个，我们就不需要计算别的 w_{ij})和相应的圈．

196

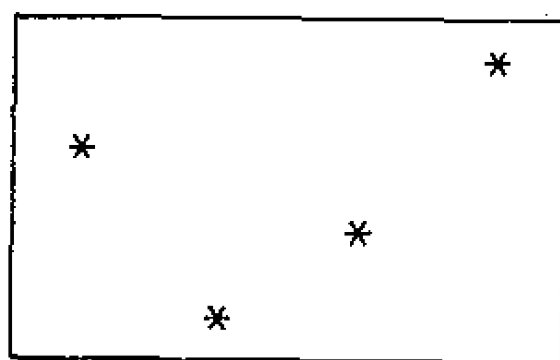
这是下一个表：

	5	6	6	5
4	$^{1}1-\varepsilon$	$^{1}(-1)$	$^{2}0$	$^{1}0+\varepsilon$
3	$^{2}(-1)\varepsilon$	$^{3}(0)$	$^{2}(-2)$	$^{3}1-\varepsilon$
3	$^{4}(2)$	$^{3}0$	$^{3}1$	$^{4}(2)$
2	$^{4}(1)$	$^{4}1$	$^{5}(1)$	$^{6}(3)$

选择位置(2，1)和 $\varepsilon=1$．然后我们又选定位置(2，4)(这次没有别的选择)．下面我们计算新势和 w_{ij}：

	5	5	5	5
4	$^{1}0$	$^{1}(0)$	$^{2}(1)$	$^{1}1$
3	$^{2}1$	$^{3}(1)$	$^{2}0$	$^{3}(1)$
2	$^{4}(1)$	$^{3}0$	$^{3}1$	$^{4}(1)$
1	$^{4}(0)$	$^{4}1$	$^{5}(1)$	$^{6}(2)$

现在所有 $w_{ij}\geqslant 0$，故此表是最优的．它给出的最优指派是



最优值是 min=10. 注意第 1 个旋转步后我们就得到了最优解，但最后一个表才证明此解是最优的．

注 18.8 有 n 个工人和工作的工作指派问题有 $n!$ 个可行解．将它写成运输问题后，这些解成了此运输问题的可行域的顶点，且此可行域是一个 $n!-1$ 维的单纯形(特别地，当 $n=2$ 时，它是区间)．对每个顶点，有 2^{n-1}种选择作为基．有 $(n^2)!$ 种方法将基变量放在标准表的左端，而且有 $(n^2)!$ 种方法将非基变量放在顶端．因此我们可以为此问题设想 $((n^2)!)^2\cdot 2^{n-1}\cdot n!$ 个标准表．幸运的是，单纯

形法不需要遍历所有的表．事实上，对运输问题而言单纯形法在实际中被证明是非常有效的． 197

练习

1～4. 求解下面的工作指派问题，表中数据是每人每个工作的费用，目标是求最小总费用．

1.

1	2	3	1	2
3	1	2	1	0
0	1	3	2	1
2	3	4	3	3
1	3	3	1	2

2.

4	2	2	2	1	2
2	3	1	2	4	4
2	0	1	1	1	4
2	3	4	3	4	3
3	2	1	2	4	1
1	2	1	2	0	1

3.

2	1	4	3	4	3	1
4	2	2	2	1	5	2
2	3	5	1	2	4	4
5	2	0	1	1	1	4
2	3	4	3	4	3	5
5	3	2	1	2	4	1
1	2	1	2	0	1	5

4.

0	2	2	4	0	1	5	1	4
2	3	1	3	4	3	2	4	1
2	1	4	3	4	3	1	2	4
4	2	4	2	2	2	1	5	2
2	3	5	1	2	4	2	4	4
5	2	2	4	0	1	1	1	4
2	3	4	3	4	3	2	4	5
5	2	4	3	2	1	2	4	1
1	2	1	2	0	1	2	4	5

5～8. 求解工作指派问题 1～4，表中数据是每个工人在每个工作上的效率，目标是求最大效率． 198

第7章 矩阵对策

19. 什么是矩阵对策

你可能听说过对策论(game theory[⊖])，这是当前很多人感兴趣的问题．在此我们只是很粗浅地介绍这方面的内容，但希望这里介绍的内容能激发你进一步学习这方面知识的欲望．

在韦伯斯特词典里，“game”这个英文词有多种意思．这个词常用的一种意思是：一种娱乐或消遣的活动，即“游戏(play)”．但是，当我们在这里谈论对策时，我们头脑中想到的意思更像是：一种包括竞赛、对垒、争斗等的情形，特别是：在指定的信息下，利益对立的各方允许选择一些行动，目标是使他们的赢得最大或损失最小．

根据局中人(player)数量的多少，对策可以分成多种类型．在单人对策中，局中人的目的是使自己的赢得(payoff[⊜])最大．但是，当有多人参与对策时，他们可能会有不同的最大化的目标(赢得)．令人惊讶的是，线性规划与一类称之为矩阵对策的对策有关．事实上，这两者之间有很深的联系．特别是，矩阵对策可以通过单纯形法求解，而线性规划也可以归结成矩阵对策．

本章我们将学习矩阵对策，即二人零和对策．零和(zero-sum)的意思是，一方所赢得的正是另一方所损失的．例如，当你赌 1 000 美元和对手下棋时，胜方所赢得的 1 000 美元正是输方腰包里所损失的，因此这是一种零和赢得．然而，并不是所有的双人对策都是零和对策．

例如，想象一对夫妻正在考虑移居纽约州布法罗或是俄亥俄州阿克伦．如果移居布法罗，丈夫可以找到年薪 60 000 美元的工作，但妻子只能找到年薪 20 000 美元的工作；如果移居阿克伦，妻子可以找到年薪 60 000 美元的工作，但丈夫只能找到年薪 20 000 美元的工作．如果呆在现在的地方不动，他们都将面临失业．他们应该怎么办？这个问题的回答超出本书的范围，但是，在这个对策中赢得之和显然不等于零．

例 19.1 本例中的小型矩阵对策经常被称为“硬币正反面”游戏或“输赢便士”(matching penny)游戏．我们称两个局中人分别为他和她，他选择正面(H)或反面(T)，而她也独立地选择正面或反面．如果他们的选择相同，他付给她 1 便

⊖ game 也译作“博弈”；Game Theory 也译作“博弈论”．——译者注

⊜ payoff 通常译作“支付”，这个译法对于付出、遭受损失的局中人来说是恰当的；但对于获得回报的局中人来说，有些书译为“赢得”或“收益”，似乎更确切，所以本书译者接受了这个译法．本书英文原文中基本上没有对这个词严格区分，具体含义容易从上下文中看出．也就是说，对于获得回报的局中人来说，当谈到其“支付”时，实际上指的是“从其他局中人那里赢得的支付”．——译者注

士；否则，她付给他 1 便士. 下面是他的赢得：

$$\begin{array}{cc} & \text{她} \\ & \begin{array}{cc} \text{H} & \text{T} \end{array} \\ \text{他} \begin{array}{c} \text{H} \\ \text{T} \end{array} & \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}. \end{array}$$

■

例 19.2 另一个对策的例子是石头、剪刀、布[⊖]. 在这个对策中，两个局中人同时选择石头、剪刀、布中的一个，通常是数一、二、三开始做手势. 赢得函数由如下对策规则定义：石头砸碎剪刀、剪刀剪破布、布包住石头，而双方如果采取的策略相同，则表示平局. 一次获胜得 1 分，一次平局得 0 分，输掉一次得 −1 分，这个对策可以表示成下面的矩阵形式，其中策略 1 表示石头，策略 2 表示剪刀，策略 3 表示布

$$A=\begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

选择行的局中人的赢得表示在上面的矩阵中. ■

一般地，一个矩阵对策由一个矩阵(payoff matrix，*赢得矩阵*)给出. 一个局中人(行局中人)选择行，而另一个局中人选择列. 矩阵中对应的元素表示列局中人对行局中人的支付(行局中人的赢得). 局中人可以是人，也可以是组织、计算机或者动物. 下棋、橄榄球赛、扑克牌的二十一点游戏等对策，都可以认为是(规模非常大的)矩阵对策.

200

例 19.3 下面矩阵表示的是行局中人有 4 种选择、列局中人有 5 种选择的矩阵对策：

$$A=\begin{bmatrix} 5 & 0 & 6 & 1 & -2 \\ 2 & 1 & 2 & 1 & 2 \\ -9 & 0 & 5 & 2 & -9 \\ -9 & -8 & 0 & 4 & 2 \end{bmatrix}.$$

■

均衡

设 $A=[a_{i,j}]$是一个 m 行 n 列的矩阵. 回忆一下，对应的矩阵对策是这样定义的：他选择第 i 行，而她独立地选择第 j 列，她付给他 $a_{i,j}$(美元).

矩阵 A 是他的赢得矩阵，而她的赢得矩阵是 $-A$. 矩阵 A 的行对应于他的(纯粹)策略(或着法)，A 的列对应于她的选择.

定义 19.4 *一个策略对(i,j)称为均衡或鞍点，如果他和她都不能通过改变自己的策略获得更大回报.*

⊖ 实际上，英文中用的词是"纸"(paper)而不是"布"，这里按中文习惯译作"布". ——译者注

换言之，$a_{i,j}$既是所在列中的最大元，又是所在行中的最小元.

作为例子，考虑例 19.3. 将每一列中的最大元标上 *，把每一行中的最小元标上$'$. 那么同时标上 * 和$'$的位置就是鞍点：

$$A=\begin{bmatrix} 5^* & 0 & 6^* & 1 & -2' \\ 2 & 1^{*\prime} & 2 & 1' & 2^* \\ -9' & 0 & 5 & 2 & -9' \\ -9' & -8 & 0 & 4^* & 2^* \end{bmatrix}.$$

此例中位置$(i,\ j)=(2,\ 2)$是唯一的鞍点，对应的赢得为 1.

可以说鞍点就是对策问题的解，因为任何一个局中人都不能通过单方面改变自己的策略而得到更好的结果. 对两个局中人来说，停在这个位置看起来是合理的.（如果不是零和对策，他们可能会联合起来做决策，然后分享获得的回报.）在本节的后面，我们将会看到每个鞍点的回报都是一样的，这个值称为对策的值.

然而，对于例 19.1 和例 19.2 来说，都不存在鞍点. 为了求解没有鞍点的矩阵对策问题，我们需要引入混合策略(mixed strategy)来扩展策略的概念. [201]

仍考虑由矩阵 $A=[a_{i,j}]$给出的一般矩阵对策问题. 他选择策略的一种谨慎的方式是：他假设她是一个完美的局中人(即她具有无限的智能和记忆能力)，能看透他大脑的思维. 所以，他这样推理：如果我选择第 1 个策略，那么她一定选择对我支付最少，即$\min\limits_j a_{1,j}$；如果我选择第 2 个策略，那么她将付给我$\min\limits_j a_{2,j}$；依此类推. 因此，最好的选择是我至少得到$\max\limits_i \min\limits_j a_{i,j}$. 这就是他的收益的下限，或称为最坏情形下的赢得.

类似地，如果她相信他是无所不知的，她一定会选择使得她的损失最多是$\min\limits_j \max\limits_i a_{i,j}$(她的损失的上限)的策略.

在生活中，

$$\max_i \min_j a_{i,j} \leqslant \min_j \max_i a_{i,j} \tag{19.5}$$

总是成立的(对任何矩阵 A). 这是因为他在最坏情形下的赢得不会好于他在最好的情形下的赢得. 下面我们将不采用矩阵 A 的对策论解释来证明(19.5).

此外，当且仅当$\max\limits_i \min\limits_j a_{i,j}=\min\limits_j \max\limits_i a_{i,j}$时，对策存在鞍点，且任何鞍点处的赢得都等于$\max\limits_i \min\limits_j a_{i,j}=\min\limits_j \max\limits_i a_{i,j}$，即对策的值.

不等式(19.5)的证明 令 $b=\max\min$，$c=\min\max$，d 是与b 处于同一行而与c 处于同一列的元素. 由于b 在同一行中是最小的，所以 $b\leqslant d$. 由于 c 在同一列中是最大的，所以$d\leqslant c$. 因此，$b\leqslant d\leqslant c$，即 $b\leqslant c$.

蕴涵关系的证明

$$\max\min = \min\max \Rightarrow \text{存在均衡}. \tag{19.6}$$

令 b，c，d 的意义同上．假设 $b=c$，则 $b=d=c$．所以 d 在同一行中是最小的，在同一列中是最大的，因此这个位置(对应的策略)是一个均衡．

最后，如果(i, j)和(i', j')是两个鞍点，则(i, j')和(i', j)也是鞍点．因此下面的定义是合理的：均衡处的策略称之为最优策略(optimal strategy)．那么，鞍点是一个策略对(他的最优策略，她的最优策略)．

蕴涵关系的证明

$$\text{存在均衡} \Rightarrow \max\min = \min\max. \tag{19.7}$$

令 d 是鞍点处的赢得，因为 d 在同一列中是最大的，所以 $d \geqslant \min\max$；同时，d 在同一行中是最小的，所以 $d \leqslant \max\min$.

202

所以 $\max\min \geqslant d \geqslant \min\max$，而(19.5)表明 $\max\min \leqslant \min\max$，因此 $\max\min = d = \min\max$.

问题 19.8 找出下面矩阵中的 max min 和 min max，并检查是否存在鞍点：

$$A = \begin{bmatrix} 5 & 0 & 6 & 1 & -2 \\ 2 & 1 & 2 & 1 & 2 \\ -9 & 2 & 5 & 2 & 0 \\ -9 & -8 & 0 & 4 & 2 \end{bmatrix}.$$

解 首先计算每一列的最大值：$\max(A)=[5, 2, 6, 4, 2]$，所以 $\min\max = \min[5, 2, 6, 4, 2]=2$．然后计算每一行的最小值：$\min(A)=[-2, 1, -9, -9]$，所以 $\max\min = \max[-2, 1, -9, -9]=1$．由于 $2\neq 1$，所以没有鞍点． ■

混合策略

下面仍然考虑由 $m\times n$ 矩阵 $A=[a_{i,j}]$给出的矩阵对策问题．

定义 19.9 *局中人的混合策略是他或她的纯策略的概率分布．*

我们把他的混合策略写成如下的列向量：

$$p = [p_1, \cdots, p_m]^{\mathrm{T}},$$

其中，对于任意的 i，$p_i \geqslant 0$，且 $p_1+\cdots+p_m=1$．他的原始策略从现在开始称为纯策略(pure strategy)，可以认为是向量中对应的元素只取 0 或 1 的混合策略，因此混合策略正是纯策略的凸组合．所以，他的纯策略是他的混合策略集合的顶点，她的纯策略也是她的混合策略集合的顶点．

我们把她的混合策略写成行向量 $q=[q_1, \cdots, q_n]$，其中对于任意的 j，$q_j \geqslant 0$，且 $q_1+\cdots+q_n=1$.

当局中人选择了他们的混合策略 p 和 q 后，那么他对应的赢得(数学期望)是

$$p^{\mathrm{T}} A q^{\mathrm{T}} = \sum_{i=1}^{m} \sum_{j=1}^{n} p_i a_{i,j} q_j,$$

而她的支付为 $p^{\mathrm{T}} A q^{\mathrm{T}}$.

在例 19.1 中，他会自然地采用混合策略 $[1/2, 1/2]^{\mathrm{T}}$. 也就是说，他可以通过投掷一枚匀称的硬币来做决策，这时无论她采用什么样的纯策略或混合策略，他的赢得是 0. 类似地(这个对策是对称的)，她也会选择混合策略 $[1/2, 1/2]$，这时无论他采用什么样的策略，她的支付也是 0. 因此，这一对混合策略是一个均衡(在混合策略的意义下).

203

一般地，他会选择至少有以下赢得的混合策略(称为最优策略)：

$$他的值 = \max_{q \in Q} \min_{p \in P} p^{\mathrm{T}} A q^{\mathrm{T}},$$

其中 P 是他的所有混合策略的集合，Q 是她的所有混合策略的集合(事实上，这里的 Q 也可以换成她的所有纯策略的集合).

类似地，她的一个合理的选择应该是至多有以下支付的混合策略(称为最优策略)：

$$她的值 = \min_{p \in P} \max_{q \in Q} p^{\mathrm{T}} A q^{\mathrm{T}}.$$

极小极大定理 *对任意的矩阵 A，*

$$他的值 = 她的值.$$

换言之，在混合策略中一定存在均衡.

求解矩阵对策的含义是：为他和她找到最优策略(即均衡)，并求出对策的值(即他的值或她的值). 在下一节中，我们将把求解矩阵对策和求解线性规划联系起来，由此可以看到，对策论中的极小极大值定理本质上就是线性规划中的对偶定理.

例 19.10 (*两指猜拳*)(P. Morris) 甲、乙两个局中人同时伸出一个或两个手指并说出一个数，希望猜中两人伸出的手指的总数(只可能是 2，3，4 之一). 如果一方猜中而另一方没有猜中，猜中方将从另一方获得 1 美元. 否则，谁也不用付钱. 玩这个游戏时，甲方有六种可能的策略：(1，2)，(1，3)，(1，4)，(2，2)，(2，3)，(2，4)，这里(1，2)的意思是伸出 1 个手指并喊出 2，以此类推. 策略(1，4)，(2，2)显然是很愚蠢的，所以可以忽略这两个策略. 乙方同样有六种相同的策略.

204

这个对策可以用如下的赢得矩阵表示：

$$\begin{array}{c|cccc} \text{甲} \backslash \text{乙} & (1,2) & (1,3) & (2,3) & (2,4) \\ \hline (1,2) & 0 & 1 & -1 & 0 \\ (1,3) & -1 & 0 & 0 & 1 \\ (2,3) & 1 & 1 & -1 & -1 \\ (2,4) & 0 & -1 & 1 & 0 \end{array}$$

在这个矩阵中，正数表示甲得到的收益，负数表示他的损失．从矩阵中可以清楚地看到，如果甲、乙两人长时间地玩这个游戏，他们都不会只采用其中的一条策略而不采用其他策略．例如，如果甲笨到只采用策略(2，3)，那么乙可以只采用策略(2，4)(除非乙比甲更笨)．玩这个游戏的一种正确的方式是随机地采用这些“纯”策略(或者其中的一部分也可)．混合策略正是描述了如何把这些纯策略混合起来使用．例如，“随机地采用这 4 种策略使得每个策略被采用的机会是 1/4”就是一种混合策略． ■

问题 19.11 验证策略($[0，1/2，1/2，0]^{T}$，$[0，1/2，1/2，0]$)是一个均衡，即$[0，1/2，1/2，0]^{T}$是甲的最优策略，$[0，1/2，1/2，0]$是乙的最优策略．计算这个对策的值．

解 将赢得矩阵增广为

$$p=[0,1/2,1/2,0]^{T} \text{ 且 } q=[0,1/2,1/2,0]),$$

$$\begin{array}{c} \\ (1,2) \\ (1,3) \\ (2,3) \\ (2,4) \\ p \end{array}\begin{array}{c} \begin{array}{ccccc} (1,2) & (1,3) & (2,3) & (2,4) & q \end{array} \\ \left[\begin{array}{ccccc} 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0^{*\prime} \end{array}\right] \end{array}.$$

易知$(p，q)$位置是一个鞍点，对应的赢得 0 就是这个对策的值． ■

这个对策是对称的，因为其赢得矩阵满足 $A=-A^{T}$，是反对称矩阵．显然，对称对策的值应该是 0，并且 s 是一个局中人的最优策略当且仅当 s^{T} 是另一个局中人的最优策略．但是，找到最优策略并不总是这么容易．

205

本章的主要目的是将矩阵对策和线性规划联系起来．下面我们对一般的对策论作几点评注，然后结束本节的讨论．

对策论引起数学家对决策过程建立模型．这一理论的开创工作通常追溯到 Cournot (1889)，他讨论了两个寡头企业(即假设市场上只有两家企业相互竞争)如何选择产品的销售价格．另一个早期的贡献来自于 Zermelo(1913)，他研究了国际象棋比赛，证明了或者白方总是能赢，或者黑方总是能赢，或者双方总是能保持平局．但是 Zermelo 没有说清楚这三种可能性中究竟哪一种是正确的，更不

用说准确地告诉双方应该如何走棋了，否则这个游戏也不会直到今天仍然是对人类和计算机的一种挑战. Zermelo 的结果是基于对这个游戏数学结构的分析，因此也适用于三子连珠(tic-tac-toe)游戏，因为这个游戏也具有这些结构性质. 对这个游戏，双方能保证平局是众所周知的.

无论决策的目的是为了获利、娱乐或其他目的，如赢得战争或赢得选举，对策论总是需要对决策过程及不同决策组合产生的结果建立数学模型进行分析.

在对策论中，决策者称为*局中人*(player)，决策称为*着法*(move)，而局中人使用着法的每一个机会称为一个*局势*(position). 着法其实是局中人以确定性或一定几率的方式改变自己在对策中的局势. 多个局中人同时使用着法的局势可以通过一系列局势进行，其中的每一个局势正好表示一个局中人的着法而不影响整个对策的策略. 因此，不失一般性，我们可以假定在每一个不是终止局势的局势，只有一个局中人使用着法.

终止局势(terminal position)是指所有局中人都不能再使用着法的局势，并且在每个终止局势都确定了一个*结果*. 在实际生活中，结果可能是一个局中人赢了而另一个局中人输了，也可能是每个局中人同时赢了或输了一些钱，或者一些并非如此清晰确定的情况，如有些局中人高兴而另一些局中人不高兴. 在对策论中，结果通常用数值型的赢得来表示局中人对结果的价值*评估*(valuation)或*偏好*(preference)，而偏好就是相对价值.

206

对一个给定的局中人来说，这个局中人面对每一个局势采取的着法的集合称为这个局中人的*纯策略*. 为局中人面对每一个局势采取何种着法指定一定概率的规则称为*行为策略*(behavioral strategy). Kuhn (1950)得到一个结论，如果局中人面对任何局势都能够记忆他或她在以前的局势所采用的所有着法，那么这个局中人的任何一个行为策略都能够用一个*混合策略*(mixed strategy)表示，即在局中人的纯策略上指定一个单一的概率分布，而不需要为每一个局势分别指定概率分布. 对策的结果可以作为局中人的纯策略的函数而计算出来. 如果考虑混合策略，只能计算结果的期望值，因为对每一次具体的对策，只有纯策略能产生精确的结果. 无论哪一种情况，对策的*赢得函数*(payoff function)实际上是一个计算规则，这个规则为每个局中人指定赢得值作为所有局中人遵循的策略的函数.

用每个局中人的策略集合和赢得函数来表示对策，而不是表示成局势和着法的集合，通常是很方便的. 因为局中人可以在对策实际开始之前选择好他们的策略，当把对策表示成策略集合和赢得函数时，实际上假设了局中人同时选择各自的策略，因此每个局中人选择策略时并不知道其他局中人做选择的任何信息. 这称为对策的*策略形式*(strategic form)或*标准形式*(normal form).

根据参与对策的局中人数量的多少、局中人得到的信息的性质、是否包括机会行为、以及赢得函数的性质，对策可以分成多种类型. 当局中人多于一人时，

对一个局中人有利的结果不一定对另一个局中人也有利. 一般来说，如果不知道对方将会如何决策，一个局中人不可能知道他或她的决策导致的结果. 在对策论中，局中人如何预测其他人的策略，这个问题令人非常着迷. 在多数对策论的分析中，总是假设所有局中人都尽量精选着法，以便使自己的赢得最大. 对很多现实生活中的决策问题，包括室内游戏(如下棋)和竞争性对策(如 Cournot 讨论的两个寡头企业的对策)等，这是一个符合实际的假设. 这个假设在什么时候有效以及为什么成立是当前对策论研究中的一个热门问题(参见 Bolton, 1998; Kurland and Byrne, 2000; Byrne and Kurland, 2000). 因为我们在此讨论的重点是线性规划，而不是心理学和行为科学，所以我们在本书中总是坚持这 207 个假设.

我们将与线性规划一起分析的对策问题称为二人零和(two-person zero-sum)对策，包括前面提到的国际象棋和三子连珠游戏，但不包括 Cournot 讨论的两个寡头企业的对策. 零和的意思是所有局中人的赢得之和总是等于零. 在零和对策中，对一个局中人有利的结果总是要损害至少一个其他局中人的利益. 在二人零和对策中，两个局中人的利益正好相反，因此这种对策有时称为*严格竞争的*，因为双方不可能通过合作而同时获得利益. 凡是以一个局中人获胜结束或者以两个局中人平局结束的二人对策，如果忽略掉对局中人获胜的重要性的考虑，都可以看成是零和对策模型. 如果每个局中人的纯策略集合是有限的，那么就可以用线性规划找到每一个局中人的策略，以保证局中人可能得到的最佳回报. 后面我们将严格地说明这一方法.

注意并非在所有的对策中着法空间都是有限的. 例如，在*定时对策*(game of timing)中，一次着法是从一个连续的区间中选择一个时刻，如经典的“手枪决斗”. 1832 年，手枪决斗夺去了年仅 20 岁的法国天才数学家伽罗瓦的生命.

练习

1～10. 在以下矩阵给出的矩阵对策中，找出 max min 和 min max，并检查是否存在鞍点. 如果不存在鞍点，尽可能为两个局中人找到最好的混合策略.

1. $A=\begin{bmatrix} 0 & 0 & 6 & 1 & -2 \\ -1 & 2 & 5 & 2 & 0 \\ -4 & -8 & 0 & 4 & 2 \end{bmatrix}$.

2. $A=\begin{bmatrix} 0 & 0 & 6 & 1 & -2 \\ 2 & 1 & 2 & 1 & 2 \\ -1 & 2 & 5 & 2 & 0 \\ -4 & -8 & 0 & 4 & 2 \end{bmatrix}$.

$$3.\ A=\begin{bmatrix} 4 & 5 & 0 & -1 & 1 & -2 \\ 0 & 2 & 1 & 2 & 1 & -2 \\ 1 & -1 & 2 & 5 & 2 & 0 \\ -4 & -9 & -8 & 0 & 4 & 2 \end{bmatrix}.$$

208

$$4.\ A=\begin{bmatrix} 4 & -4 & 5 & 0 & 0 & -1 & 1 & -2 \\ 5 & 0 & 2 & 1 & 2 & 1 & 0 & -2 \\ -4 & 0 & -2 & 1 & 5 & 1 & 6 & 2 \\ 1 & -1 & 2 & 5 & 3 & 7 & 2 & 0 \\ -4 & -9 & -8 & 0 & 4 & 2 & 0 & -3 \end{bmatrix}.$$

$$5.\ A=\begin{bmatrix} 4 & -4 & 3 & 0 & 0 & 0 & -1 & 1 & -2 \\ -1 & 0 & 2 & 1 & -2 & -2 & 1 & 0 & -2 \\ -4 & 0 & -2 & -2 & 1 & -1 & 1 & 6 & 2 \\ 1 & 2 & 2 & 5 & 3 & 3 & 7 & 2 & 0 \\ -4 & -9 & -8 & 0 & 4 & 2 & 2 & 0 & 3 \end{bmatrix}.$$

$$6.\ A=\begin{bmatrix} 4 & -1 & 4 & -1 & 5 & 0 & 0 & -1 & 1 & -2 \\ -1 & 0 & 2 & 1 & 2 & 1 & 2 & 1 & 0 & -2 \\ -4 & 0 & -2 & 1 & 0 & 1 & 6 & 2 & 6 & -2 \\ 1 & -1 & 2 & 5 & 3 & 5 & 3 & 1 & 2 & 0 \\ -4 & -9 & -8 & 0 & 4 & 0 & 4 & 2 & -6 & -3 \end{bmatrix}.$$

$$7.\ A=\begin{bmatrix} 7 & -1 & 3 & -1 & 5 & 0 & 0 & -1 & 1 & -2 \\ -1 & 0 & 0 & 1 & 2 & 1 & 2 & 1 & 0 & -2 \\ -4 & 0 & -2 & 1 & 0 & 1 & 1 & 2 & 6 & -2 \\ 0 & 0 & -2 & 1 & 0 & 1 & 1 & 2 & 0 & -2 \\ 1 & -1 & 2 & 0 & 3 & 5 & 3 & 1 & 2 & 0 \\ -4 & -1 & -8 & 0 & 4 & 0 & 4 & -2 & -6 & 0 \end{bmatrix}.$$

8. $A=\begin{bmatrix} 8 & -1 & 4 & -1 & 5 & 0 & 0 & -1 & 1 & -2 \\ -1 & 0 & 2 & 1 & 2 & 1 & 2 & 1 & 0 & -2 \\ -4 & 0 & -2 & 1 & 0 & 1 & 6 & 2 & 6 & -2 \\ -4 & 8 & -2 & 1 & 0 & 1 & 8 & 2 & 6 & -2 \\ 1 & -1 & 2 & 5 & 3 & 5 & 3 & 1 & 2 & 0 \\ -4 & -9 & -8 & 0 & 4 & 0 & 4 & 0 & -6 & 3 \end{bmatrix}$.

9. $A=\begin{bmatrix} 9 & 0 & -4 & 0 & 2 & 1 & 0 & 1 & 9 & 1 & 2 & 6 & -2 \\ -4 & 3 & -4 & 0 & -2 & 1 & 9 & 1 & 9 & 1 & 2 & 6 & 0 \\ 0 & 0 & 2 & 0 & -2 & 1 & 0 & 1 & 9 & 6 & 2 & -6 & -2 \\ -4 & 0 & -4 & 0 & -2 & 1 & 0 & 1 & 9 & 6 & 2 & 6 & 3 \\ 0 & 0 & 0 & 0 & -2 & 1 & 5 & -2 & 9 & 6 & 2 & 6 & 0 \\ 4 & 0 & -4 & 0 & -2 & 1 & 0 & 1 & 9 & 6 & -2 & 6 & -2 \\ -4 & 0 & 9 & 0 & -2 & 3 & 0 & 1 & 9 & 6 & 2 & 0 & -2 \end{bmatrix}$.

209

10. $A=\begin{bmatrix} 0 & -1 & 0 & 2 & 1 & 2 & 1 & 2 & 1 & 0 & -2 \\ -1 & 1 & 0 & -2 & 1 & 0 & 1 & 6 & 2 & 6 & -2 \\ -4 & 0 & 0 & -2 & 1 & 0 & 1 & 6 & 2 & 6 & -2 \\ 0 & -4 & 0 & -2 & 1 & 1 & 1 & 6 & 2 & 6 & -2 \\ 0 & -4 & 0 & -2 & 1 & 0 & 1 & 6 & 2 & 6 & -2 \\ 2 & 0 & 0 & -2 & 1 & 0 & 1 & 1 & 2 & 6 & -2 \\ -4 & 0 & -2 & 1 & 0 & 1 & 0 & -1 & 2 & 6 & -2 \\ 1 & -1 & 2 & 5 & 3 & 1 & 0 & 3 & 1 & 2 & 0 \\ -4 & -5 & -8 & 0 & 4 & 0 & 4 & 20 & 0 & -6 & -3 \end{bmatrix}$.

11. 证明：如果(i, j)和(i', j')是两个鞍点，则(i, j')和(i', j)也是鞍点.

12. Blotto 上校（P. Morris）

甲方军队有 3 个师的兵力，正在进攻由乙方军队 4 个师防守的一个城镇. 进入该城镇的道路有两条，双方军队的兵力必须分布在两条道路上. 如果进攻者在某条路上的兵力超过防守者，那么进攻者将战胜防守者并占领该城镇. 这个城镇的价值是两个师. 如果防守者在某条路上的兵力超过进攻者，那么进攻者将损失掉相应的兵力. 甲方的策略有：(3，0)，(2，1)，(1，2)和(0，3)，其中数字如(2，1)的意思是第一条道路分配两个师，第二条道路上分配一个师. 乙方的策略

有五个：(4，0)，(3，1)，(2，2)，(1，3)和(0，4)．赢得矩阵为

$$\begin{array}{cc} A\backslash B & (4,0) \quad (3,1) \quad (2,2) \quad (1,3) \quad (0,4) \\ \begin{array}{c}(3,0)\\(2,1)\\(1,2)\\(0,3)\end{array} & \begin{bmatrix} -3 & 0 & 4 & 3 & 2 \\ 0 & -2 & -1 & 2 & 1 \\ 1 & 2 & -1 & -2 & 0 \\ 2 & 3 & 4 & 0 & -3 \end{bmatrix} \end{array}$$

从甲方的观点看，策略(2，1)和(1，2)似乎是最合理的，因为潜在的损失最多是两个师，这样的选择使其潜在的利益最大．有趣的是，如果乙方知道甲方会采用策略(2，1)或(1，2)，他将总是采用策略(2，2)，并总是能赢得这次战役．

检验这个对策的值是 1/4，甲方的最优策略是

$$[9/68,23/68,28/68,8/68]^{\mathrm{T}},$$

乙方的最优策略是

$$[5/12,0,2/12,0,5/12].$$

210

20. 矩阵对策与线性规划

用线性规划求解矩阵对策

虽然我们已经学习过如何建立和求解线性规划，但将矩阵对策归结为线性规划仍是一个具有挑战性的问题．在标准形式下，他具有 m 个策略而她具有 n 个策略的二人零和对策可以用 $m\times n$ 的矩阵 $A=[a_{ij}]$表示，其中 a_{ij} 是他采用策略 i 而她采用策略 j 时他的赢得．我们将使用行和列代替策略(例如，他采用第 i 行而她采用第 j 列)．在二人零和对策中，我们不需要指定她的赢得，因为双方的赢得之和为 0，所以她的赢得正是他的赢得的负值．因此，对于零和对策，习惯上总是简单地只说赢得，意思就是他的赢得．因此，我们说他希望使自己的赢得最大，而她想使自己的支付最小．

我们将采用混合策略，并试图为每个局中人找到最好的混合策略．作为纯策略上的一种概率分布，总是假设他和她的策略是相互独立的．我们将他的混合策略表示成 $p=[p_1，\cdots，p_m]^{\mathrm{T}}$，而她的混合策略表示成 $q=[q_1，\cdots，q_n]$．由于分布是相互独立的，概率论基本知识告诉我们，如果他采用 p 而她采用 q，则他对应赢得的数学期望是 $p^{\mathrm{T}}Aq^{\mathrm{T}}$．对于混合策略 p，乘积 $p^{\mathrm{T}}A$ 是 A 的行线性组合，其中第 i 行以他采用这行的概率 p_i 进行了加权．因此，乘积 $p^{\mathrm{T}}A$ 结果的第 j 列(只有一个元素)是她采用第 j 列时的期望支付．将 $p^{\mathrm{T}}A$ 的第 j 列按照 q_j 进行加权并求和，就得到了总的期望赢得．所以，他希望最大化 $p^{\mathrm{T}}Aq^{\mathrm{T}}$，而她希望最小化 $p^{\mathrm{T}}Aq^{\mathrm{T}}$．

在例 19.2 石头、剪刀、布的游戏中，如果他使用混合策略 $p=[1/2，1/2，0]^{\mathrm{T}}$(一半时间出石头，一半时间出剪刀)，那么她将采用策略 $q=[1，0，0]$(总是出石头)，期望从他那里赢得 1/2．如果她采用策略 $q=[1/3，2/3，0]$，期望的赢得将是

1/6. 这通过计算 $p^{T}A=[-1/2, 1/2, 0]$容易看出来. 如果她采用混合策略 $q=[1/4, 1/4, 1/2]$ (1/4 时间出*石头*, 1/4 时间出*剪刀*, 一半时间出*布*), 那么他将采用策略 $p=[0, 1, 0]^{T}$(总是出*剪刀*), 期望从她那里赢得 1/4. 如果他采用策略 $p=[1/3, 2/3, 0]^{T}$, 期望的赢得将是 1/12. 这通过计算 $Aq^{T}=[-1/4, 1/4, 0]^{T}$ 容易看出来.

211

现在的问题是, 在不能控制 q 的情况下, 他如何使 $p^{T}Aq^{T}$ 最大? 类似地, 在不能控制 p 的情况下, 她如何使 $p^{T}Aq^{T}$ 最小? 以前我们讨论的线性规划问题中只有一个决策者, 而不是两个. 这正是二人零和对策严格竞争特性的关键所在, 每个局中人都可以预测对手的行为, 确定自己赢得值的上下限, 从而求解这个对策问题. 进一步假设每个局中人都希望优化自己的赢得, 二人零和对策中就有如下问题: 不管他或她怎么做, 另一方总是希望使对手得分尽量低. 所以, 他们从一个给定的策略中期望得到的赢得总是采用该策略时他们可能得到的最坏的赢得. 这种技巧(用最坏情形下的赢得来衡量每个策略的好坏)使得另一方的选择总可以从赢得函数中推导出来! 因此, 他可以用 $f(p)=\min p^{T}A$ 来衡量每个策略 p 的好坏, 其中极小是对 $p^{T}A$ 的所有元素取极小. 这个道理和$\min\limits_{q} p^{T}Aq^{T}$ 是一样的, 其中极小是对她的所有混合策略 q 取的.

所以, 他所面对的问题是

$$\max f(p) = \min p^{T}A$$

而她所面对的问题是

$$\min g(q) = \max Aq^{T}.$$

需要强调的是, 并非任何向量 p 和 q 都是允许的, 只有概率分布才是策略. 回忆石头、剪刀、布的游戏中, 我们看到

$$f([1/2, 1/2, 0]^{T}) = -1/2, \quad g([1/4, 1/4, 1/2]) = 1/4.$$

换言之, 如果他使用策略 $p=[1/2, 1/2, 0]^{T}$, 可以保证他的赢得至少是 $-1/2$. 这是一个可以得到保证的最小值, 也就是他希望最大化的一个下界. 对于她来说, 如果使用策略 $q=[1/4, 1/4, 1/2]$, 她可以保证她的支付至多是 1/4. 这是一个可以得到保证的最大值, 也就是她希望最小化的一个上界. 显然, 他希望赢得的任何一个下界, 也是她支付的任何一个上界的下界. 类似地, 她的支付的任何一个上界, 也是他希望赢得的任何一个下界的上界. 也就是说, $g(q)\geqslant f(p)$. 在石头、剪刀、布的游戏中, 如果他使用混合策略 $p^{T}=[1/3, 1/3, 1/3]$, 则 $p^{T}A=[0, 0, 0]$, 所以无论她采用什么策略, 期望的结果都是平局(即 $f(p)=0$).

212

类似地, 如果她使用混合策略 $q=[1/3, 1/3, 1/3]$, 则 $Aq^{T}=[0, 0, 0]^{T}$, 所以无论他采用什么策略, 期望的结果也是平局(即 $g(q)=0$). 所以, $f([1/3, 1/3, 1/3]^{T})=0=g([1/3, 1/3, 1/3])$. 由于他有一个策略可以保证他的赢得至少是 0, 而她有一个策略可以保证她的支付至多是 0, 所以任何一

方都不能做得更好．因此，在这个对策中，$p^T=q=[1/3, 1/3, 1/3]$对两个局中人来说都是最优的，每个局中人的策略保证各自的最佳支付，因为对手的能力限制了支付的上下界．

用这种方式我们已经解决了一个方面的复杂性，即每个局中人的目标函数与另一个局中人的策略的相关性．但仍然存在另一个方面的复杂性，即局中人新的目标函数不是线性的．在讨论这种复杂性之前，还有另外两点需要说明．首先，任意一个 $m\times n$ 的矩阵 $A=[a_{ij}]$可以看成一个二人零和对策，其中他有 m 个纯策略而她有 n 个纯策略，且当他采用第 i 行而她采用第 j 列时，赢得是 a_{ij}．至于一个矩阵是否正好表示了现实世界中的一个对策，则是一个建模问题，而与作为对策问题来求解的数学结构无关．二人零和对策经常被称为*矩阵对策*，正是基于这种同一性，所以具有实值赢得函数和有限多个纯策略的二人零和对策与所有有限维实矩阵构成一一对应．

需要说明的另一点是，在石头、剪刀、布的对策例子中，对所有纯策略 p 和 q 来说，$-1=f(p)<g(q)=1$．但这个不等式关系并不总是成立的(参见例 19.3)．

历史评注

1. 在任何时刻，所有局中人都知道精确局势的对策，称为完美信息(perfect information)下的对策．下棋、三子连珠、二十一点等都是完美信息下的对策，但掷硬币猜正反面，石头、剪刀、布等则不是．对策的展开形式是一种称之为树的图形结构，其中的节点表示局势，而从每个非终止节点指向其他节点的弧表示着法．如果以这种方式定义的展开形式是有限的，则这个对策称为有限的．Zermelo(1913)证明了如下结论：任何一个完美信息下的有限对策，都存在纯策略的均衡．这个均衡可以通过从终止局势方向回溯到所有其他局势而得到(动态规划)． 213

矩阵中的鞍点是对应二人对策中的一对最优纯策略，所以 Zermelo 的定理表明，任何一个完美信息下的二人零和有限对策的标准形式，都存在鞍点．

2. Nash (1949)证明了对于有限多个局中人，每个局中人的纯策略也有限的任何一个对策，一定存在混合策略的均衡．均衡(现在被广泛地称为纳什均衡)，实际上是策略的集合，即为每个局中人指定一个策略，其他局中人按照自己的策略行动时，此策略是该局中人的最佳响应策略．换言之，任何一个局中人都不可能通过单方面地改变其策略而改善其赢得．纳什(Nash)所讨论的对策问题的特征可以刻画为有限标准形式(finite normal form)．(从理论上看，可以为展开形式的无限分支定义赢得，只是此时可能永远无法到达终止节点，所以有些对策的展开形式是无限的，但标准形式是有限的．)纳什证明了对于有限标准形式的对策，一定存在混合策略的均衡，其中纯策略看作是混合策略的一种

特殊情形.

矩阵对策中的鞍点实际上是纯策略意义下的均衡，而一个矩阵对策不存在鞍点时，我们找到的最优混合策略也一定是纳什均衡. 纳什在他的证明中使用了拓扑学中很有威力的不动点定理. 在学会如何用线性规划求解最优策略后，我们就可以推导出二人零和对策这种特殊情况下的纳什定理，这只需要利用四种可能性(即任何线性规划的四种可能的有限形式)的对偶定理就可以了.

现在我们回到为任意一个 $m\times n$ 的矩阵 A 表示的矩阵对策寻找最优混合策略的问题. 前面我们已经成功地将每个局中人的目标函数表示成该局中人自己的策略函数，而不再依赖于对手的策略. 然而，作为希望最小化和最大化的函数，$f(p)$ 和 $g(q)$不是线性的！为了处理这种复杂性，我们需要利用另一个技巧. 实际中为了把一个问题表示成线性规划或差分方程或其他数学形式，通常需要使用多种技巧. 如果希望将来用应用数学方法解决实际问题，应该努力钻研本书中的这类技巧，经过较长一段时间后，你将成为发明这类技巧的老手. 在我们这里所讨论的问题中，所用的技巧是为行局中人引入一个变量 λ，替代非线性目标函数 $f(p)$；同时，为列局中人引入一个变量 μ，替代非线性目标函数 $g(q)$.

214

我们使用的技巧要想成功，就必须保证最大化 λ 等价于最大化 $f(p)$，最小化 μ 等价于最小化 $g(q)$. 这可以通过增加关于 λ 和 $f(p)$的关系的约束，以及增加关于 μ 和 $g(q)$的关系的约束做到. 回忆一下，$f(p)=\min\{p^{\mathrm{T}}Aq^{\mathrm{T}} \mid q$ 是她的策略$\}$，这里的条件“q 是她的混合策略”在数学上可以用以下 $m+1$ 个线性约束表示：

$$q_j \geqslant 0(\text{对所有 } j),\text{其中 } q=[q_1,\cdots,q_n],\text{且}\sum_{i=1}^{n} q_i = 1.$$

这些条件将向量 q 限定为纯策略有限集合上的概率分布，我们可以利用这些约束条件用 λ 表示 $f(p)$. 乘以 q 相当于离散情形下对概率分布 q 的积分，结果正是 $p^{\mathrm{T}}A$ 的分量的线性组合. 并且由于对 q 有上面的约束条件，所以这种线性组合是一种凸组合，因此最后的结果一定位于 $p^{\mathrm{T}}A$ 的分量形成的凸包(convex hull)中. 按照通常的说法，我们是对 $p^{\mathrm{T}}A$ 的分量计算加权平均值(weighted average). 直观的想法是，n 个数的平均值(即使是加权平均值也一样)，一定以这些数中的最大数为上界，而以这些数中的最小数为下界.

现在我们就可以将为他或她寻求最佳策略的问题建模成线性规划. 他的线性规划问题是

$$\lambda \to \max, p^{\mathrm{T}}A-[\lambda,\cdots,\lambda] \geqslant 0, \sum_{i=1}^{m} p_i = 1, p \geqslant 0,$$

而她的线性规划问题是

$$\mu \to \min, Aq^{\mathrm{T}}-[\mu,\cdots,\mu]^{\mathrm{T}} \leqslant 0, \sum_{j=1}^{n} q_j = 1, q \geqslant 0.$$

为了将这两个规划问题放到标准的单纯形表中，我们必须把上面的表达式改写成第 2 章中所要求的形式，这里将它写成列单纯形表的形式. 我们增加的变量 λ 并不能总是限定为非负值，这个问题可以有多种处理方法. 标准的处理技巧是变量替换，即令 $\lambda=\lambda'-\lambda''$，其中变量 λ' 和 λ'' 就可以限定为只取非负值(参见第 2 章). 类似地，可以令 $\mu=\mu'-\mu''$. 此外，用 I 表示由 m 个 1 组成的一行，用 J 表示由 n 个 1 组成的一列. [215]

现在，我们可以将上面的两个线性规划问题写成如下的标准单纯形表：

$$\begin{array}{c} \\ -p \\ -\lambda' \\ -\lambda'' \\ 1 \\ \\ \\ \end{array}\begin{array}{c} \begin{array}{cccc} q & \mu' & \mu'' & 1 \end{array} \\ \left[\begin{array}{cccc} -A & I^{\mathrm{T}} & -I^{\mathrm{T}} & 0 \\ J^{\mathrm{T}} & 0 & 0 & -1 \\ -J^{\mathrm{T}} & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{array}\right] \\ \begin{array}{cccc} \| & \| & \| & \| \\ * & * & * & \lambda\to\max \end{array} \end{array}\begin{array}{l} \\ =*\geqslant 0 \\ =*\geqslant 0 \\ =*\geqslant 0 \\ =\mu\to\min \\ \\ \\ \end{array}$$

注意在上面的表格中，她的问题对应于行问题，而他的问题对应于列问题，这两个问题互为对偶. 由于两个问题都有可行解(例如，取 $p=I^{\mathrm{T}}/m$，$q=J^{\mathrm{T}}/n$)，对偶定理可以保证 $\min(f)=\max(g)$，也就是说他的最优值＝她的最优值. 因此，极小极大值定理可以直接从对偶定理得到.

当 $v=\min(f)=\max(g)>0$ 时(这可以将矩阵 A 的所有元素增加一个常数 v_0 使得矩阵的所有元素为正数做到，但在这种修改后的问题求解完成以后，不要忘记减去 v_0 才是原问题的解)，可以假设 $\lambda>0$，并令 $x=p/\lambda$. 此时，他的问题为如下形式：

$$\text{minimize } 1/\lambda = Ix \text{ subject to } x\geqslant 0, x^{\mathrm{T}}A\geqslant J^{\mathrm{T}}.$$

而她的问题是

$$1/\mu = yI \to \max, Ay^{\mathrm{T}}\leqslant I, y\geqslant 0,$$

其中 $y=q/\mu$. 这两个问题都可以写成如下的标准单纯形表：

$$\begin{array}{c} \\ -x \\ 1 \\ \\ \\ \end{array}\begin{array}{c} \begin{array}{cc} y & 1 \end{array} \\ \left[\begin{array}{cc} -A & I^{\mathrm{T}} \\ -J^{\mathrm{T}} & 0 \end{array}\right] \\ \begin{array}{cc} \| & \| \\ * & -1/\lambda\to\max \end{array} \end{array}\begin{array}{l} \\ =*\geqslant 0 \\ =-1/\mu\to\min \\ \\ \\ \end{array}$$

这个问题解决之后，我们可以计算对策值 $v=1/Ix$，他的最优混合策略是 $p=xv$，她的最优混合策略是 $q=yv$. [216]

这种技巧在单纯形表中节省了两行和两列，如果我们用手工方法进行计算是有意义的. 此外，另一个好处是，这个单纯形表是行可行的，所以不必执行单纯形法的第一阶段.

例 20.1 求解矩阵对策

$$\begin{bmatrix} 1 & 3 & 5 \\ 4 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix}.$$

这个对策问题可以利用优超关系(优超关系使我们可以去掉最后一行和最后一列)和图解法(见第 21 节)进行求解. 在此我们还是说明如何使用单纯形方法. 此前，我们首先找到行局中人至少得到 1 的行和列局中人至少付出 3 的列. 所以这个对策的值位于 1 到 3 之间，但不存在鞍点. 为了找到均衡，我们必须使用混合策略 $p=[p_1, p_2, p_3]^T$，$q=[q_1, q_2, q_3]$. 由于这个对策的值是正的，所以在标准单纯形表中省去两行和两列：

$$\begin{array}{cc} & \begin{array}{cccc} q_1/\mu & q_2/\mu & q_3/\mu & 1 \end{array} \\ \begin{array}{c} -p_1/\lambda \\ -p_2/\lambda \\ -p_3/\lambda \\ 1 \end{array} & \begin{bmatrix} -1 & -3 & -5 & 1 \\ -4 & 0 & -1 & 1 \\ -1 & -2 & 0 & 1 \\ -1 & -1 & -1 & 0 \end{bmatrix} \begin{array}{l} =v_1 \\ =v_2 \\ =v_3 \\ =-1/\mu \to \min \end{array} \\ & \begin{array}{cccc} =u_1 & =u_2 & =u_3 & =-1/\lambda \quad \to \max \end{array} \end{array}$$

先以(v_1, q_1/μ)位置为旋转元进行旋转，再以(v_2, q_2/μ)位置为旋转元进行旋转，就可以得到最优的单纯形表. 最优解是 $q_1/\mu=q_2/\mu=1/4$，$q_3/\mu=0$，$-1/\mu=-1/2=\min$，$p_1/\lambda=1/3$，$p_2/\lambda=1/6$，$p_3/\lambda=0$，$-1/\lambda=-1/2=\max$. 所以，最优策略是 $p=[p_1, p_2, p_3]=[2/3, 1/3, 0]$，$q=[q_1, q_2, q_3]=[1/2, 1/2, 0]$，对策的值是 2. ■

例 20.2 求解如下赢得矩阵给出的矩阵对策：

$$A=\begin{bmatrix} 0 & 2 & -1 & 0 \\ -2 & 0 & 4 & -3 \\ 1 & -4 & 0 & 2 \\ 0 & 3 & -2 & 0 \end{bmatrix}.$$

217

由于矩阵 $A=-A^T$ 是反对称的，所以这个对策是对称的，其值是 0. 但是，这对寻找最优策略并没有什么帮助，然而我们知道对于列局中人来说，她的最优策略是满足 $Aq^T\leqslant 0$(即 $qA\geqslant 0$)的混合策略.

这个对策不存在鞍点，也不存在任何优超关系. 为了节省行和列，我们将赢得矩阵的每个元素增加 1，使得对策的值变成正的(即 1)，写出标准的行单纯形表：

$$\begin{array}{c} \begin{matrix} q_1 & q_2 & q_3 & q_4 & 1 \end{matrix} \\ \begin{bmatrix} -1 & -3 & 0 & -1 & 1 \\ 1 & -1 & -5 & 2 & 1 \\ -2 & 3 & -1 & -3 & 1 \\ -1 & -4 & 1 & -1 & 1 \\ -1 & -1 & -1 & -1 & 0 \end{bmatrix} \begin{matrix} = v_1 \\ = v_2 \\ = v_3 \\ = v_4 \\ \to \min \end{matrix} \end{array}$$

通过三次旋转变换，可以用 v_1，v_2，v_3 换出 q_1，q_2，q_3，从而得到最优单纯形表：

$$\begin{array}{c} \begin{matrix} v_1 & v_2 & v_3 & q_4 & 1 \end{matrix} \\ \begin{bmatrix} * & * & * & * & 4/7 \\ * & * & * & * & 1/7 \\ * & * & * & * & 2/7 \\ * & * & * & * & * \\ * & * & * & * & -1 \end{bmatrix} \begin{matrix} = q_1 \\ = q_2 \\ = q_3 \\ = v_4 \\ \to \min \end{matrix} \end{array}$$

所以，列局中人的最优策略是[4/7，1/7，2/7，0]，行局中人的最优策略是[4/7，1/7，2/7，0]^{T}. ■

线性规划归结为矩阵对策

现在我们将任意的对偶线性规划对(13.4)

$$\begin{array}{cc} & \begin{matrix} x & 1 \end{matrix} \\ \begin{matrix} -y \\ 1 \end{matrix} & \begin{bmatrix} A & b \\ c & d \end{bmatrix} \begin{matrix} = u \\ = z \to \min \end{matrix} \quad \begin{matrix} x \geqslant 0, u \geqslant 0 \\ y \geqslant 0, v \geqslant 0 \end{matrix} \\ & \begin{matrix} = v & = w & \to \max \end{matrix} \end{array}$$

归结为矩阵对策. 初看起来，这似乎是不可能的，因为线性规划不一定是可行的，而每一个矩阵对策一定有最优解.

218

我们可以进一步考虑如下的矩阵：

$$M = \begin{bmatrix} 0 & -A & -b \\ A^{\mathrm{T}} & 0 & -c^{\mathrm{T}} \\ b^{\mathrm{T}} & c & 0 \end{bmatrix}.$$

这个矩阵的规模是$(m+n+1)\times(m+n+1)$，其中 $m\times n$ 是矩阵 A 的规模. 由于 $-M=M^{\mathrm{T}}$，所以对应的矩阵对策是对称的，对策的值是 0. 事实上，假设无论她怎么做，他总能至少有一定的赢得 $\varepsilon>0$，即对于任意的列向量 p，$p^{\mathrm{T}}M\geqslant\varepsilon>0$. 那么，她也可以使用同样的策略 $q=p^{\mathrm{T}}$ 来对付他，无论他怎么做，她也总能至少有一定的赢得 ε. 当双方都使用这个策略时，赢得 $p^{\mathrm{T}}Aq^{\mathrm{T}}$ 应该同时既是正的又是负的，这相互矛盾.

假设 $x=\bar{x}$，$u=\bar{u}$ 是行问题的最优解，$y=\bar{y}$，$v=\bar{v}$ 是列问题的最优解. 令 e 是 $\bar{x}$ 和 $\bar{y}$ 的所有元素的和再加上 1.

这时，$\bar{p}=[\bar{y}^{\mathrm{T}}，\bar{x}，1]^{\mathrm{T}}/e$ 是他的一个混合策略，对应的赢得为

$$\bar{p}^{\mathrm{T}}M = [\bar{y}^{\mathrm{T}}, \bar{x}, 1]M/e$$
$$= [\bar{x}A^{\mathrm{T}} + b^{\mathrm{T}}, -\bar{y}^{\mathrm{T}}A + c, -\bar{y}^{\mathrm{T}}b - \bar{x}c^{\mathrm{T}}]/e = [\bar{u}^{\mathrm{T}}, \bar{v}, 0]/e \geqslant 0$$

(其中用到：对于最优解一定有 $z=w$)因此，$\bar{p}$ 是他的最优策略. 注意列向量 $\bar{p}$ 的最后一个元素是 $1/e$，不是 0.

反过来，给定他的任何一个最优策略 $p=\bar{p}$(也就是说 $\bar{p}$ 也是她的最优策略)，假定 $\bar{p}$ 的最后一个元素不是 0(计算为 $1/e$)，我们可以写成 $\bar{p}^{\mathrm{T}}=[\bar{y}^{\mathrm{T}}, \bar{x}, 1]^{\mathrm{T}}/e$，其中 $\bar{y}^{\mathrm{T}}$ 和 $\bar{x}$ 是相应规模的非负行向量. 由于 $\bar{p}$ 是最优的，所以 $\bar{p}^{\mathrm{T}}M \geqslant 0$. 因此，

$$\bar{x}A^{\mathrm{T}} + b^{\mathrm{T}} \geqslant 0, -\bar{y}^{\mathrm{T}}A + c \geqslant 0, -\bar{y}^{\mathrm{T}}b - \bar{x}c^{\mathrm{T}} \geqslant 0.$$

这说明 $x=\bar{x}$，$u=A\bar{x}^{\mathrm{T}}+b$ 是前面所介绍的行问题的最优解，$y=\bar{y}$，$v=c-\bar{y}^{\mathrm{T}}A$ 是列问题的最优解(再次用到 $\min(z)=\max(w)$).

因此，在对策问题的最优策略(最后一个元素非零)和以上两个线性规划的最优解之间，存在一一对应关系. 如果不存在最后一个元素非零的最优策略，那么相应的线性规划问题也没有最优解.

219

练习

1～4. 求解以下矩阵对策.

1. $\begin{bmatrix} 1 & 3 & 5 \\ 4 & 0 & -1 \\ -1 & 2 & 0 \end{bmatrix}$.

2. $\begin{bmatrix} 1 & -3 & 4 \\ 1 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix}$.

3. $\begin{bmatrix} 1 & -3 & 5 & 1 & -3 & 5 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 2 & 0 & 1 & 2 & 0 \end{bmatrix}$.

4. $\begin{bmatrix} 1 & -3 & 5 & 1 & -3 & 5 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 2 & 0 & 1 & 2 & 0 \end{bmatrix}$.

5～10. 求解第 19 节练习中的 1～6 题.

11～13. 给定如下线性规划(其中所有 $x_i \geqslant 0$)，写出一个与这个线性规划及其对偶问题等价的矩阵对策问题.

11. $x_1 + 2x_2 + x_3 + x_4 + x_5 + 3x_6 \to \min$,

$3x_1 + x_2 + x_3 + 2x_4 + x_5 + x_6 \geqslant 5$,

$3x_1 + x_2 + x_3 + 2x_4 + x_5 + x_6 \geqslant 4$,

$3x_1 + x_2 + x_3 + 2x_4 + x_5 - x_6 = 3$.

12. $x_1 + 2x_2 + 4x_3 + x_4 + x_5 + x_6 \to \max$,

$3x_1 + x_2 + x_3 + 2x_4 + x_5 - 2x_6 + x_7 \geqslant 5$,

$3x_1 + x_2 + x_3 + 2x_4 + x_5 + x_6 - x_7 \geqslant 6$,

$3x_1 + x_2 + x_3 + 2x_4 + x_5 - x_6 \leqslant 7$.

13. $x_1 + 2x_2 + x_3 + x_4 + x_5 + 3x_6 + x_7 + x_8 \to \min$,

$3x_1 + x_2 + x_3 + 2x_4 + x_5 + x_6 + x_7 - 3x_8 \geqslant 1$,

$3x_1 + x_2 + x_3 + 2x_4 - x_5 + x_6 + x_7 + 3x_8 \geqslant 5$,

$3x_1 + x_2 + x_3 + 2x_4 + x_5 - x_6 + x_7 + x_8 \geqslant 1$.

220

21. 其他方法

当矩阵对策中的某个局中人只有唯一的策略时，这个矩阵对策很容易求解：另一个局中人只需要根据对手的唯一策略选择使自己的赢得最大. 实际上，这种情形下的矩阵对策退化成了单人对策. 这样一个例子是娱乐场所的"二十一点"赌博游戏，因为庄家(发牌人)的策略是固定的.

当矩阵对策中的某个局中人只有两个纯策略时(即赢得矩阵只有两行或两列)，这个矩阵对策也不需要用单纯形法求解，此时使用图解法就可以很容易地求解这个矩阵对策，下面将用两个例子进行说明. 有时，我们可以去掉一些冗余的策略，降低赢得矩阵的规模，使得一些规模相当大的对策问题容易求解. 另一种简单的想法是检查是否存在鞍点，这种想法有时也很有效. 或者能够找到一个鞍点，问题已经解决；或者能够找到对策值的上界和下界.

优超与图解法

例 21.1 考虑如下矩阵给出的矩阵对策：

$$A = \begin{bmatrix} -1 & 6 & 1 & 6 \\ 9 & 2 & 3 & 1 \\ 8 & -1 & 3 & 0 \end{bmatrix}.$$

如果手上没有计算机，我们如何求解这个对策问题呢？注意，他的第二个策略 r2 总是比第三个策略 r3 给出更多的或者相等的赢得. 换言之，

$$\text{第二行} \geqslant \text{第三行}.$$

我们称策略 r2 *优超*(dominate) r3.

给定他的任何一个混合策略

$$p=[p_1,p_2,p_3]^{\mathrm{T}}=p_1\mathrm{r}1+p_2\mathrm{r}2+p_3\mathrm{r}3$$

如果他在 p 中用第二个策略 r2 代替 r3，不会损失更多. 特别地，他一定存在一个不使用 r3(对应的元素为 0)的最优策略，我们可以去掉矩阵的第三行而不会引起对策值的改变. 所以，我们只需要求解下面这个规模更小的对策问题：

221

$$A=\begin{bmatrix}-1 & 6 & 1 & 6\\ 9 & 2 & 3 & 1\end{bmatrix}.$$

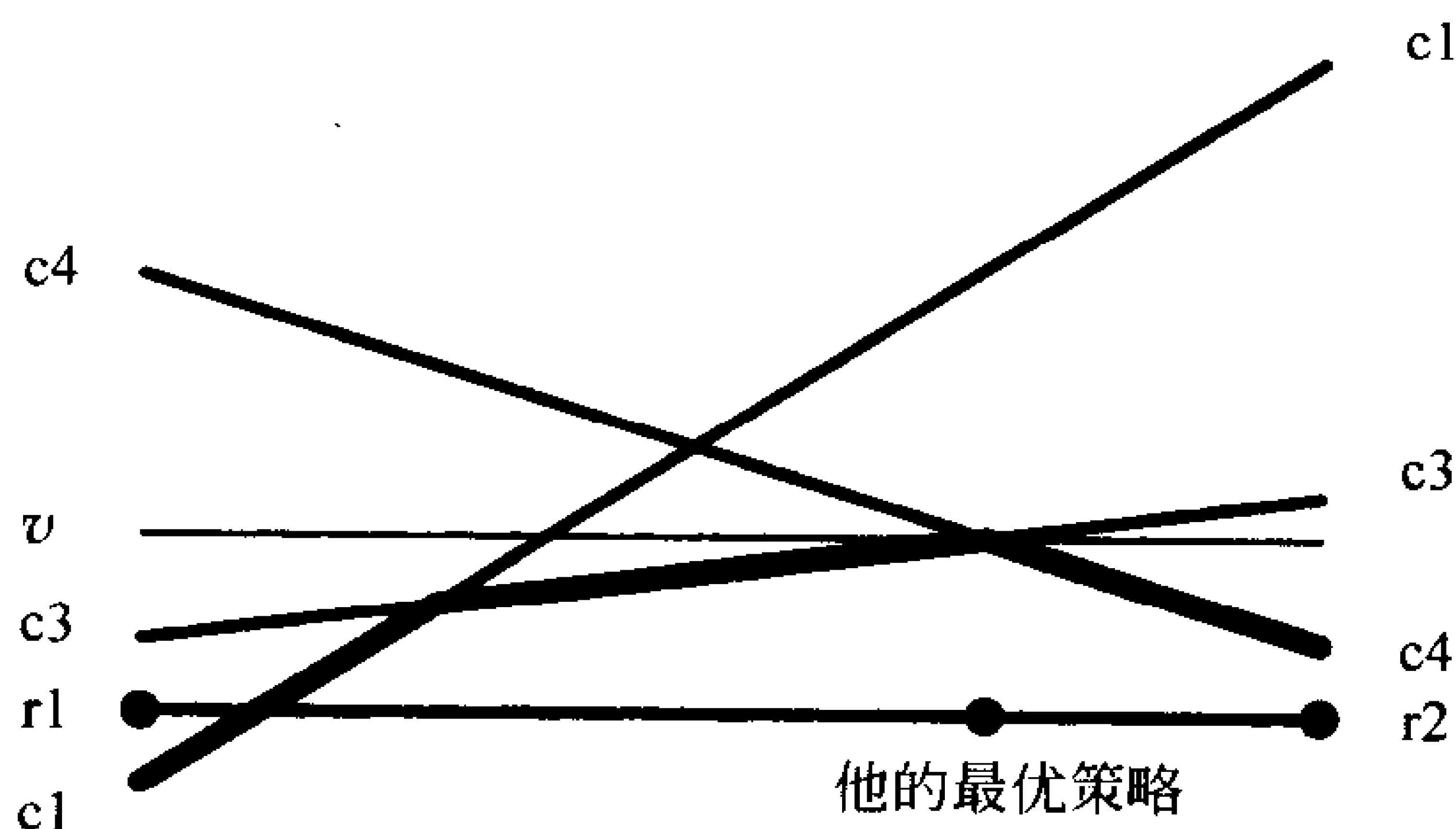
其次，注意到她的第四个策略 c4 优超她的第二个策略 c2：

第四列≤第二列.

所以我们可以去掉矩阵的第二列而不会改变对策的值，得到规模更小的对策问题：

$$A=\begin{bmatrix}-1 & 1 & 6\\ 9 & 3 & 1\end{bmatrix}.$$

现在，在行和列中都不存在优超关系. 由于他只有两个纯策略(r1，r2)，我们可以用图解法求解这个矩阵对策. 也就是说，在水平轴上画出一个单位区间，表示他的混合策略，她的策略 c1，c3，c4 用这个区间上的线性函数表示. 这些函数的最小点组成一个上凸函数，其最大值是 $v=\min\max$.



所以，他的最优策略是$(1-p_2)\mathrm{r}1+p_2\mathrm{r}2$，其中

$$1\cdot(1-p_2)+3p_2=6(1-p_2)+1\cdot p_2=v,$$

于是 $7p_2=5$. 因此，他的最优策略是(2/7)r1+(5/7)r2，对策的值是 $v=17/7$.

她的最优策略 $q_3\mathrm{c}3+(1-q_3)\mathrm{c}4$ 是 c3 和 c4 的组合，在我们的图形中给出的是一个常数函数 17/7：

$$1\cdot q_3+6(1-q_3)=17/7=3q_3+1\cdot(1-q_3),$$

所以 $q_3=5/7$. 因此，她的最优策略是(5/7)c3+(2/7)c4.

222

现在我们可以给出对于原来大矩阵问题的最后答案：

对策的值是 17/7；

他的最优策略是 $p=(2/7)\mathrm{r}1+(5/7)\mathrm{r}2=[2/7,\ 5/7,\ 0]^{\mathrm{T}}$；

她的最优策略是 $q=(5/7)\mathrm{c}3+(2/7)\mathrm{c}4=[0,\ 0,\ 5/7,\ 2/7]$. ■

回忆一下，他是行局中人，她是列局中人，赢得矩阵是对他给出的.

例 21.2 假设我们要求解如下矩阵给出的矩阵对策：

$$\begin{bmatrix} -1 & 2 & 2 & 0 \\ 2 & 0 & -1 & 2 \\ 1 & 1 & 1 & 1 \\ 0 & 2 & 1.5 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

首先将每一列中的最大元素标上 *，将每一行中的最小元素标上“′”：

$$\begin{bmatrix} -1' & 2^* & 2^* & 0 \\ 2^* & 0 & -1' & 2^* \\ 1' & 1' & 1' & 1' \\ 0' & 2^* & 1.5 & 0' \\ 0' & 0' & 1 & 2^* \end{bmatrix}.$$

因为没有任何位置被标记两次，所以不存在鞍点. 我们有

$$\max\min = 1 \leqslant \text{对策的值} \leqslant \min\max = 2.$$

我们称行局中人为“他”，他的纯策略为 r1－r5，列局中人为“她”，她的纯策略为 c1－c4 ：

$$\begin{array}{c} \\ r1 \\ r2 \\ r3 \\ r4 \\ r5 \end{array} \begin{array}{c} \begin{array}{cccc} c1 & c2 & c3 & c4 \end{array} \\ \begin{bmatrix} -1 & 2 & 2 & 0 \\ 2 & 0 & -1 & 2 \\ 1 & 1 & 1 & 1 \\ 0 & 2 & 1.5 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \end{array}.$$

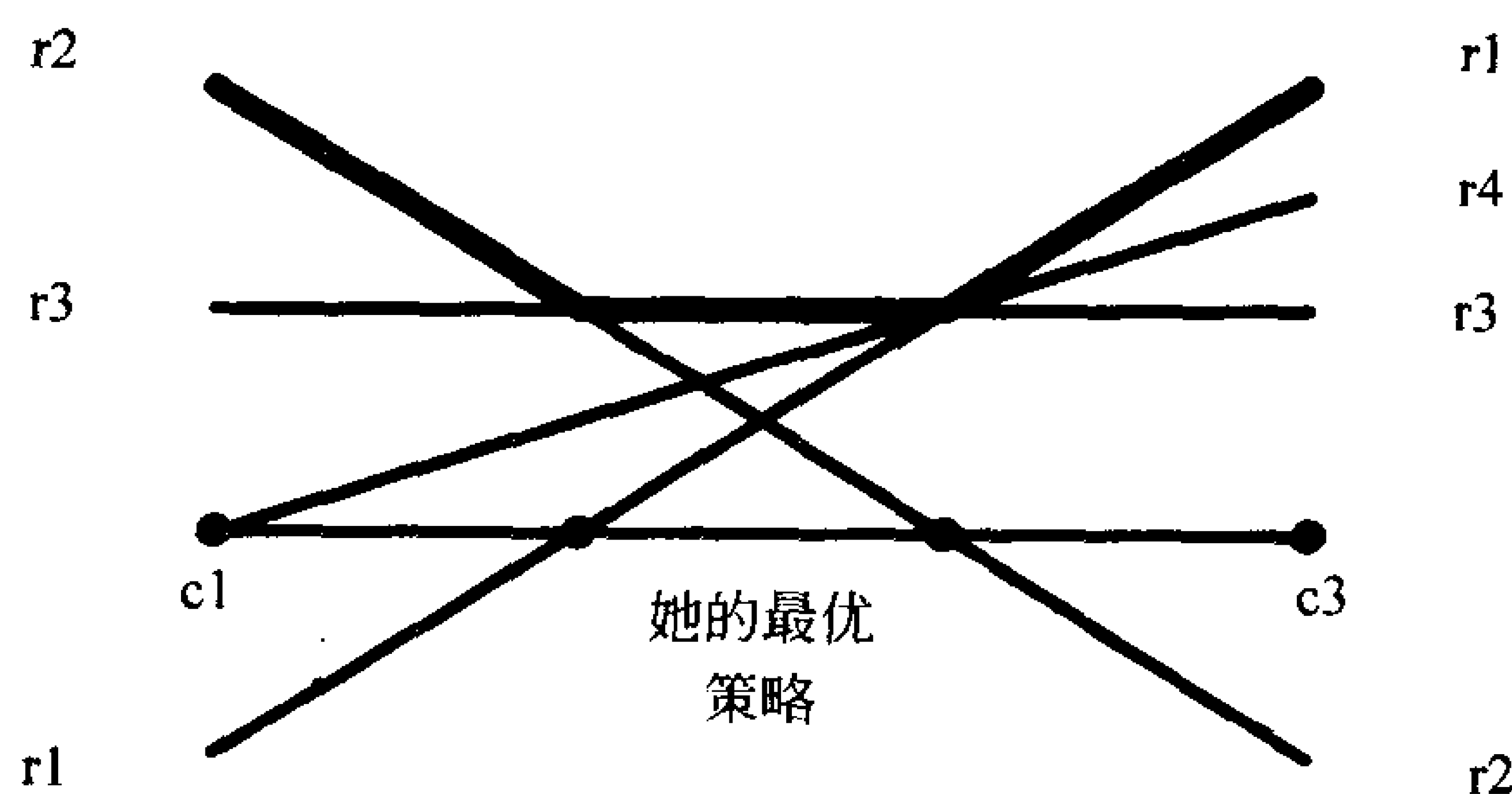
根据优超关系，可以去掉 c4(与 c1 比较)，然后去掉 r5(去掉 c4 后与 r3 比较)，再去掉 c2(去掉 r5 后与 c3 比较). 此时剩下 4×2 矩阵：

$$\begin{array}{c} \\ r1 \\ r2 \\ r3 \\ r4 \end{array} \begin{array}{c} \begin{array}{cc} c1 & c3 \end{array} \\ \begin{bmatrix} -1 & 2 \\ 2 & -1 \\ 1 & 1 \\ 0 & 1.5 \end{bmatrix} \end{array}.$$

223

现在我们作图，在水平轴上连接 c1 和 c3 的线段表示她的混合策略，而他的策略用这条线段上的函数表示，即对应的赢得.

由于她总是计算她的最坏情形下的赢得，所以我们首先对她的选择计算最大值. 然后我们在这些最大值中取最小值，这是一个分段线性函数. 显然，最小值是 1，在[2/3，1/3](图中左端点)和[1/3，2/3]（右端点)中的任何一点都可以达到这个值. 他的最优策略是 r3，图中用一条水平线表示. 现在我们可以给出对



于原来的 5×4 矩阵问题的最后答案：对策的值是 1，他的最优策略是 r3(即混合策略为$[0, 0, 1, 0, 0]^T$)；她的最优策略是(1/3)c1＋(2/3)c3＝[1/3，0，2/3，0]. 她的所有最优策略是

$$a\mathrm{c}1 + (1-a)\mathrm{c}3 \text{ 其中 } 1/3 \leqslant a \leqslant 2/3.$$

为了再次检验这个答案是否正确，我们用她的最优策略 q＝[1/3，0，2/3，0]和 q'＝[2/3，0，1/3，0]增广赢得矩阵，确实得到均衡：

$$\begin{array}{c} \\ \mathrm{r}1 \\ \mathrm{r}2 \\ \mathrm{r}3 \\ \mathrm{r}4 \\ \mathrm{r}5 \end{array}\begin{array}{c} \begin{array}{cccccc} \mathrm{c}1 & \mathrm{c}2 & \mathrm{c}3 & \mathrm{c}4 & q & q' \end{array} \\ \begin{bmatrix} -1 & 2 & 2 & 0 & 1 & 0 \\ 2 & 0 & -1 & 2 & 2/3 & 1/3 \\ 1 & 1 & 1 & 1 & 1^{*\prime} & 1^{*\prime} \\ 0 & 2 & 1.5 & 0 & 2/3 & 1/3 \\ 0 & 0 & 1 & 2 & 0 & 0 \end{bmatrix} \end{array}.$$

■

下面我们给出一些其他技巧，如变尺度、变换、以及对称性等，这些技巧可能有用. 如果我们将矩阵 A 的每个元素乘以同一个正数 t，那么最优策略仍然不变，新对策的值是原来对策的值的 t 倍.

224

如果我们将矩阵 A 的每个元素加上同一个正数 t，那么最优策略仍然不变，新对策的值是原来对策的值加上 t.

如果我们交换行或者列，对策的值不变.

例如，下面赢得矩阵给出的矩阵对策的值是 2：

$$\begin{bmatrix} 3 & 2 & 0 \\ 2 & 1 & 3 \\ 1 & 4 & 2 \end{bmatrix}.$$

为了明白这一点，我们将矩阵的每个元素加上－2，然后交换前两列，此时得到一个反对称矩阵.

虚构对策(Brown 方法)

局中人从某种策略 $p^{(0)}$，$q^{(0)}$开始，然后他找到了对付她的策略 $q^{(0)}$ 的最优响应策略 $p^{(1)}$；她也找到了对付他的策略 $p^{(0)}$ 的最优响应策略 $q^{(1)}$. 下一步他找到了对付她的策略$(q^{(0)}+q^{(1)})/2$ 的最优响应策略 $p^{(2)}$；她也找到了对付他的策略

$(p^{(0)}+p^{(1)})/2$ 的最优响应策略 $q^{(2)}$.

在第 t 步，他找到了对付她的如下策略的最优响应策略 $p^{(t)}$：

$$\bar{q}^{(t-1)}=(q^{(0)}+\cdots+q^{(t-1)})/t,$$

她也找到了对付他的如下策略的最优响应策略 $q^{(t)}$：

$$\bar{p}^{(t-1)}=(p^{(0)}+\cdots+p^{(t-1)})/t.$$

可以看出，我们只使用先前的策略 $\bar{q}^{(t-1)}$、当前的最优响应策略 $q^{(t)}$ 和时间 t 来计算她的累积策略：

$$\bar{q}^{(t)}=\bar{q}^{(t-1)}(1-1/t)+q^{(t)}/t.$$

同样可以计算他的累积策略. 为了使最优响应策略唯一，可以将所有可能的最优响应策略按照相同的权重加权混合起来.

可以证明(由 J. Robinson 证明)，对于任意的矩阵 A，这种方法(由 Brown 提出)都是可行的. 也就是说，当 t 趋于无穷时，平均赢得值 $\min(\bar{p}^{(t)\mathrm{T}}A)$ 和 $\max(A\bar{q}^{(t)\mathrm{T}})$ 都收敛到对策的值. 序列 $\bar{p}^{(t)}$ 的每个极限点都是他的一个最优策略，序列 $\bar{q}^{(t)}$ 的每个极限点都是她的一个最优策略. [225]

由于每一个线性规划都可以归结为一个矩阵对策，所以 Brown 方法也能用于求解任何一个线性规划问题. 还有很多其他求解矩阵对策和线性规划的迭代方法. 例如，对对策论做出主要贡献的 J. von Neumann 建议对对称对策使用连续时间形式的 Brown 方法. 最新的方法之一是 Karmarkar 方法，这个方法比 Brown 方法更复杂，但收敛更快. 关于内点法的更多信息，参见附录.

问题 21.3 用虚拟对策迭代法，求解如下赢得矩阵(行局中人的收益，单位是美元)给出的矩阵对策：

$$A=\begin{bmatrix}1&-1&0&2&-3&1\\1&1&-1&0&1&1\\-1&0&1&-2&2&1\end{bmatrix}.$$

解 由于没有给出初始点，我们可以随意选择一个，以两个局中人的最佳纯策略作为初始点. 我们称列局中人为 Ann，她的纯策略是 c1，c2，c3，c4，c5，c6；称行局中人为 Bob，他的纯策略是 r1，r2，r3. Ann 计算每一列中的最大值，并标上 *：max=[1，1，1，2，2，1]，所以她的最佳纯策略是 c1，c2，c3，c6. 如果 Ann 选择这些策略组成的任何混合策略，她付给 Bob 最多 min max=1 美元.

Bob 计算每一行中的最小值，并标上“′”：min=[−3，−1，−2]^{T}，所以他的最佳纯策略是 r2. 如果 Bob 选择这个策略，他最少得到 max min = −1 美元. 由于 max min 和 min max 不相等，所以不存在鞍点(没有任何位置被标记两次)，但我们知道对策的值 v 一定位于 −1 和 1 之间.

由于 Ann 不知道应该采取哪个纯策略作为初始策略，所以她采用如下混合策略：

$$q^{(0)} = [1,1,1,0,0,1]/4.$$

相应最坏情形下的赢得是

$$-u_0 = -\min(Aq^{(0)T}) = -\min([1/4,1/4,1/4]^T) = -1/4.$$

所以，在最坏情形下，她支付给他 1/4，因此我们知道 $-1 \leqslant v \leqslant 1/4$. 他选择 $p^{(0)} = \text{r}2 = [0, 1, 0]^T$. 现在进行一次迭代，并用 $A.q$ 表示 Aq^T，用 $p.A$ 表示 p^TA：

226

$p^{(0)}$	$= [0,1,0]^T$	$q^{(0)}$	$= [1,1,1,0,0,1]/4$
$A.q^{(0)}$	$= [1,2,1]^T/4$	$p^{(0)}.A$	$= [1,1,-1,0,1,1]$
$\min(p^{(0)}.A) = -1$		$\max(A.q^{(0)}) = 1/4$	
$p^{(1)}$	$= [1,1,1]^T/3$	$q^{(1)}$	$= [0,0,1,0,0,0]$
$\bar{p}^{(1)}$	$= [0,1,0]^T$	$\bar{q}^{(1)}$	$= [1,1,5,0,0,1]/8$
$A.\bar{q}^{(1)}$	$= [1,-3,5]^T/8$	$\bar{p}^{(1)}.A$	$= [0,1,-1,0,1,2]/2$
$\min(\bar{p}^{(1)}.A) = -1/2$		$\max(A.\bar{q}^{(1)}) = 5/8.$	

于是，Bob 有一个策略 $\bar{p}^{(1)}$ 使他至多付出 1/2，而 Ann 有一个策略 $q^{(0)}$ 使她至多付出 1/4，对策值的上下界之间仍然有差异：$-1/2 \leqslant v \leqslant 1/4$. 用手工方法继续这个迭代过程是很费时间的. 如果猜不出最优策略(对这个具体例子中的问题来说这是很容易的)，那就尝试手工计算一下练习中的第 5 题. ■

练习

1～4. 求解以下矩阵对策.

1. $\begin{bmatrix} 3 & 2 & 0 \\ 3 & 1 & 3 \\ 4 & 4 & 2 \end{bmatrix}$.

2. $\begin{bmatrix} 3 & 2 & 0 \\ 3 & 1 & 0 \\ 0 & 4 & 2 \end{bmatrix}$.

3. $\begin{bmatrix} 3 & 2 & 0 \\ 3 & 1 & 4 \\ 3 & 2 & 1 \\ 4 & 4 & 2 \end{bmatrix}$.

4. $\begin{bmatrix} 3 & 2 & 0 \\ 3 & 1 & 3 \\ 0 & 0 & 2 \\ 3 & 2 & 2 \\ 3 & 3 & 1 \\ 5 & 4 & 0 \end{bmatrix}$.

5. 求解问题 21.3 的矩阵对策.

6～14. 求以下矩阵对策的值. 如果不能求出精确值，尽可能算出对策值最好的上界和下界.

6. $\begin{bmatrix} 0 & 9 \\ 5 & 2 \end{bmatrix}$.

7. $\begin{bmatrix} 0 & 1 & 2 & 3 \\ -1 & 0 & 4 & 5 \\ -2 & -4 & 0 & 6 \\ -3 & -5 & -6 & 0 \end{bmatrix}$.

227

8. $\begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ 2 & 2 & 1 & 2 & 2 \\ 1 & 0 & 0 & 2 & 2 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 2 & 1 \end{bmatrix}$.

9. $\begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 9 & -1 & 8 \\ 5 & 2 & 3 & -2 \end{bmatrix}$.

10. $\begin{bmatrix} 1 & 2 & 0 & 3 \\ 5 & 7 & 5 & 6 \\ 4 & 5 & 1 & 6 \\ 7 & 8 & 3 & 9 \\ 0 & 1 & 5 & 1 \end{bmatrix}$.

11. $\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 1 & 2 \\ -1 & -1 & 0 & 1 & 3 \\ -1 & -1 & 0 & 0 & 4 \end{bmatrix}$.

12. $\begin{bmatrix} 0 & 1 & -1 & 1 & 1 \\ -1 & 0 & 1 & 1 & 2 \\ 2 & -1 & 0 & 1 & 3 \\ -1 & 3 & 1 & 1 & 2 \\ -1 & -1 & 0 & 0 & 4 \end{bmatrix}$.

13. $\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 & 1 & 1 & -2 \\ -1 & -1 & 0 & 2 & 0 & 1 & 3 \\ -1 & 0 & 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & -1 & 0 & 0 & 4 \end{bmatrix}$.

14. $\begin{bmatrix} 4 & 4 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 4 & 2 & 1 & 1 & -2 \\ 0 & 1 & 4 & -1 & 2 & 1 & 1 & -2 \\ -1 & -1 & -1 & 2 & 0 & 4 & 1 & 3 \\ -1 & 0 & 3 & 1 & 4 & 1 & 1 & 2 \\ 0 & 0 & 2 & -1 & 0 & 4 & 0 & 4 \end{bmatrix}$.

第8章 线性逼近

22. 什么是线性逼近

在使用数学方法求解实际问题之前，我们经常需要收集一些数值型的数据. 这并不是一件容易的事情. 例如，如何衡量一个人高兴的程度？人的身高一般认为是一个没有争议的量，然而精确的测量结果显示，即使对于成年人，在同一天中的身高也不是常数. 那么光速又如何呢？它是一个物理常数，一个熟练的观测者利用相同的工具，并尽可能消除掉各种可能引起变化的因素后，所得到的多次观测结果难道还会不完全相同吗？当然不同！即使最仔细地进行实验，也会得到不同的结果.

在一架飞机上数一下乘客的数量，即使这样简单的事情有时也会出现偏差而引起航班的延误. 此外，航空公司不仅希望知道当前航班上的乘客数量，还希望知道未来航班上的乘客数量！(参见第 24 节习题 4.)

科学家使用的典型方法是，多次观测这些量，然后取平均数. 平均数是用一个单一的值来综合表示一系列数据，它总是位于最小值和最大值之间. 最经常使用的平均数有均值、中位数和中程数. 观测值与选用的平均数之差称为*残差*(residual)，或称为*偏差*(discrepancy)、*纵向差*(vertical deviation)、*误差项*(error term)等. 纵向差这个名称来源于图形上平均数用一条水平的直线表示.

现在我们定义这些平均数，并分别解释它们在什么意义下是最优的. 考虑 m 个观测值(数)，它们的算术均值定义为

$$(a_1+\cdots+a_m)/m.$$

229

为了定义其他平均数，把观测值按升序排列比较方便：$a_1\leqslant\cdots\leqslant a_m$. 特别地，$a_1=\min(a_i)$，$a_m=\max(a_i)$. 这样，中程数定义为

$$(a_1+a_m)/2.$$

当 m 是奇数时，中位数定义为 $a_{(m+1)/2}$；当 m 是偶数时，中位数定义为满足 $a_{m/2}\leqslant x\leqslant a_{m/2+1}$ 的任何数 x. 在有些教科书中，为了使得定义的中位数唯一，而将中位数定义为 $(a_{m/2}+a_{m/2+1})/2$，本书则将这种定义称为样本中位数(sample median)或中值(central value).

例 对于数 2，5，5，7，均值为 19/4=4.75，中程数为 9/2=4.5，中位数为 5. ■

注 计算中程数和中位数时，并不一定要将所给的数进行排序. 可以通过 $m-1$ 次比较找到 $\min(a_i)$，然后通过 $m-2$ 次比较找到 $\max(a_i)$. 所以计算中程数只需要经过 $2m-3$ 次比较和 2 次算术运算. 此外，中位数

也可以在线性时间内计算出来(参见附录). 然而, 如果要把 m 个数进行排序, 则至少需要 $\log_2 m! \geqslant m(\log m-1)$ 次比较.

这三个平均数之所以被经常使用, 是因为在下面的三种意义下它们分别是最优的(给定数据的最佳拟合).

定理 22.1 均值 x_2 是下面的优化问题的最优解:

$$\sum_{i=1}^{m}(a_i-x)^2 \to \min.$$

证明 目标函数可以写成

$$\sum_{i=1}^{m}(a_i-x)^2 = m(x-x_2)^2+C,$$

其中 C 是常数. 显然, 在 $x=x_2$ 时上面式子取到最小值 C, 且最优解是唯一的. ■

所以均值就是最小二乘拟合(least square fit), 或称为最佳 l^2 拟合.

定理 22.2 中程数 x_∞ 是下面优化问题的最优解:

230

$$\max_i |a_i-x| \to \min.$$

证明 如前所述, 将所给的数排好序, 则目标函数可以写成

$$\max(a_1-x, x-a_1, a_m-x, x-a_m)$$

$$=\begin{cases} a_m-x & \text{若 } x \leqslant x_\infty \\ x-a_1 & \text{否则} \end{cases} = |x-x_\infty| + (a_m-a_1)/2.$$

显然, 在 $x=x_\infty$ 时上面式子取到最小值 $\min=(a_m-a_1)/2$, 且这个最优解是唯一的. ■

所以中程数可以称为最佳 l^∞ 拟合.

定理 22.3 一个数 x_1 是一个中位数, 当且仅当它是下面优化问题的最优解:

$$\sum_{i=1}^{m}|a_i-x| \to \min.$$

证明 如前所述, 将所给的数排好序, 则在下面的两个无穷区间和 $m-1$ 个闭区间的每个区间上, 目标函数都是仿射函数:

$$x \leqslant a_1; a_i \leqslant x \leqslant a_{i+1} (i=1,2,\cdots,m-1); x \geqslant a_m.$$

其斜率分别是 $-m$, $-m+2i$ (假定 $a_i \neq a_{i+1}$), m. 显然, 当 m 是偶数且 $a_{m/2} \neq a_{m/2+1}$ 时, 最优解集合是区间 $a_{m/2} \leqslant x_1 \leqslant a_{m/2+1}$. 否则只有一个最优解, 即

$$x_1 = \begin{cases} a_{m/2} = a_{m/2+1} & m \text{ 为偶数且 } a_{m/2} = a_{m/2+1} \\ a_{(m+1)/2} & m \text{ 为奇数.} \end{cases}$$

这与中位数的定义是一致的. ■

利用取整函数$\lfloor \ \rfloor$，中位数 x_1 的定义可以写成一个公式：

$$a_{\lfloor (m+1)/2 \rfloor} \leqslant x_1 \leqslant a_{\lfloor (m+2)/2 \rfloor}.$$

定理 22.3 告诉我们中位数正是最佳 l^1 拟合.

注 类似地，我们可以定义最佳 l^p 拟合，但 $p=1$，2，∞是最常用的. 之所以如此，其中的一个原因是这些拟合比较容易计算. 231

为了介绍双变量模型，我们先看一个例子. 你的体重是超重、过轻还是正好？也许你会说“这是我自己个人的事情，我不想谈论这件事”. 一些专家对体重超重和过轻都提出了警告，但正常的体重应该是多少？

对这个问题有不同的看法. 有些人说，一个人理想的体重与基因类型、肥胖细胞的数量、健康状况、生活方式、个人爱好有关，而与其他人的体重没有关系.

但在一些情形下，一个人的体重与一群人的平均体重进行比较是很重要的. 例如，一个人如果希望能够成为橄榄球队进攻线上的队员，可能就需要比赛场上其他队员的体重要重些. 相扑运动员可能也希望比他的对手体重重些. 如果生活在寒冷的环境里，或者参加耐久力测试比赛(如流行的“生存”电视表演)，可能体重也要重些. 另一方面，职业赛马骑师却可能希望自己是所有选手中最轻的一个，一些径赛运动员也可能会努力减轻自己的体重.

判断你体重状况的简单的但未必是正确的一种方法，是把你的体重与其他人(与你同等的人)的平均体重进行比较. 但是你使用的平均值不同，选择比较的同伴不同，答案可能也不同. 不过，专家建议使用一种更复杂的方法：将你的体重与你自己的人体参数(如身高)进行比较.

例如，网址 http://health. yahoo. com/health 建议采用下面的方法：

确定你体重的一个简单方法是采用下面的公式：

女人：基础身高 5 英尺算 100 磅，此后每增加 1 英寸增加 5 磅. 按照这个公式就可以计算出合理的体重.

男人：基础身高 5 英尺算 106 磅，此后每增加 1 英寸增加 6 磅.

用 w 表示体重，h 表示高度，这个公式可以表示成 $w=5h-200$ (对女人)和 $w=6h-254$ (对男人). 令人疑惑的是，这里的系数 5，-200，6，-254 从哪里来的？是某个伟大的科学家在象牙塔里运用自然界的基本规律计算出来的，还是某个人根据现实生活中的数据进行统计分析而得到的？如果是后一种情况，那么当你使用这些公式时，实际上是隐蔽地将你的体重与其他人的体重进行比较. 一种明显的方式是：假设理想的体重是一个常数 b(即公式是 $w=b$)，并通过一组同伴的平均数来计算这个常数. 回顾一下，我们前面考虑过三种不同的平均数概念. 232

虽然你可能希望自己是橄榄球队里体重最重的一个，但你的医生关心的却是你的身高和体重的关系. 所以我们现在讨论身高和体重的关系，将会用到更复杂一点的数学. 我们的目标是如何确定模型 $w=a+bh$ 中的系数 a 和 b. 假设你是选修线性规划课程的 49 个学生之一，并且班上所有的同学都愿意公开有关自己身体特征的参数. 考虑每个同学的身高 h_i 和体重 w_i，$i=1, 2, \cdots, 49$. 我们可以在平面上画出这些点(h_i, w_i)，然后寻找这个散点图的图式. 可能这些点正好集中在一条直线的周围，此时我们希望找到直线 $w=a+bh$ 来最好地拟合这些数据.

直线 $w=a+bh$ 正好通过所有的 49 个点不太可能，换言之，由 49 个线性方程 $w_i=a+bh_i(i=1, 2, \cdots, 49)$组成的方程组关于两个未知数 a 和 b 不太可能有解. 我们只好找近似"解". 自然，我们希望找到最好的近似解，但如何比较两个不同的近似解才能决定哪个解更好呢?

换言之，需要对两个变量的 49 个目标函数 $|w_i-a-bh_i|$ 最小化，所以我们希望将它们组合成一个目标函数，使我们构造的优化问题有意义.

有很多方法可以做到这一点. 最常用的方法是：

$$e_1^2+e_2^2+\cdots+e_{49}^2 \to \min, \tag{22.4}$$

$$|e_1|+|e_2|+\cdots+|e_{49}| \to \min, \tag{22.5}$$

$$\max(|e_i|) \to \min, \tag{22.6}$$

其中 $e_i=w_i-a-bh_i$ 称为*残差*、*纵向差*或*误差项*. 以(22.4)、(22.5)、(22.6)为目标函数，就可以对 $p=1, 2, \infty$分别得到最佳的 l^p 拟合. [233]

注 目标函数 $\max|e_i|$ 被认为是 $p=\infty$情形的原因如下：当 $p\to\infty$时，

$$\|e\|_p=\left(\sum_{i=1}^{m}|e_i|^p\right)^{1/p} \to \max(|e_i|).$$

一种更复杂的身高和体重的关系式是 $w=a+bh+ch^2+dh^3$. 令 $e_i=w_i-(a+bh_i+ch_i^2+dh_i^3)$，就可以极小化三个目标函数(22.4)、(22.5)、(22.6)中的任意一个. 此外，除了身高，其他参数也可以纳入模型中. 例如，美国海军所采用的利用人体各部分周长的方法是：对男人，测量身高、颈围和腰围；对女人，测量身高、腰围、颈围和臀围.

身高 h 和体重 w 的关系有一个简单的模型：$w=ch^2$. 使用这个模型的组织有：CDC（疾病控制与预防中心，美国联邦政府中负责公民健康与安全保护的领导机构），NIH(国家健康研究院，另一个联邦政府机构)，AHA（美国心脏病协会). 当身高 h 以米为单位，体重 w 以千克为单位时，比值 $w/h^2(\mathrm{kg/m^2})$称为体重指数(body mass index，BMI).

根据 CDC、NIH 和 AHA 的观点，对于预测人体健康所面临的危险来说，利用体重指数 BMI 比利用单一的体重更有价值. AHA 曾经认为 BMI 位于 19

至 25 之间时是健康的，但 2001 年这个区间被改成了 18.5 至 24(最新的修改参见 AHA 的网页 http：//www.americanheart.org，其他网页上也会随着时间的变化给出类似的数据.)在一份研究报告中(1996 年)，研究人员发现有 49%的美国女人、59%的美国男人的体重指数 BMI 大于 25，也就是说超过 50% 的美国人体重超重. 在 50 岁到 60 岁之间的人群中，64%的女人、73%的男人超重.

下面是 CDC 对“成年人的健康与 BMI 有什么关系”这个问题的回答：健康成年人的 BMI 应该位于 18.5 和 24.9 之间，这个区间是基于体重对疾病和死亡的影响得出的. 1998 年，NIH 也将这个区间作为“正常体重”的范围.

1998 年 NIH 的指南指出，BMI 也有不足之处(例如对于正在长身体的人)，同时建议对于 BMI 位于 25 至 34.9 之间的人，除了使用 BMI 外，还需要考察腰围.

234

所以，我们可以画出点(h_i^2，w_i)，然后尝试用一条通过原点的直线拟合它们. 令 $e_i=w_i-ch_i^2$，当选择(22.4)、(22.5)、(22.6)中的一个作为目标函数后，我们就得到了一个关于单变量 c 的优化问题.

注　不应该仅仅根据大学教科书上介绍的内容就对自己的健康状况作出评判.

问题 22.7　根据下面给出的数据，对 $p=1$，2，∞分别求出最佳 l^p 拟合 $w=ch^2$：

i	1	2	3
身高 h(m)	1.6	1.5	1.7
体重 w(kg)	65	60	70

对于同一个 p，与采用模型 $w/h^2=c$ 的最佳拟合的 c 进行比较；对于同一个 p，与模型$w=b$的最佳拟合的最小值进行比较.

解　$p=1$ 的情形：可以将这个问题转化成 4 个变量的线性规划问题，然后用单纯形法求解(参见下面的第 23 节). 也可以直接考虑如下目标函数的非线性优化问题：

$$f(c)=|65-1.6^2c|+|60-1.5^2c|+|70-1.7^2c|$$

这是一个没有约束条件的极小化问题. 函数 $f(c)$是一个分段仿射的凸函数，在以下几个点斜率会发生变化：

$$c=70/1.7^2\approx 24, c=65/1.6^2\approx 25, c=60/1.5^2\approx 27.$$

函数 $f(c)$的斜率是

$$-1.6^2-1.5^2-1.7^2<0,\text{当 } c\leqslant 70/1.7^2,$$

$$-1.6^2-1.5^2+1.7^2\approx-2,\text{当 } 70/1.7^2\leqslant c\leqslant 65/1.6^2,$$

$$1.6^2-1.5^2+1.7^2\approx 3,\text{当 } 65/1.6^2\leqslant c\leqslant 60/1.5^2,$$

$$1.6^2+1.5^2+1.7^2>0,\text{当 } c\geqslant 60/1.5^2.$$

易见，函数 $f(c)$ 的极小点是：

235

$$c=x_1=65/1.6^2\approx 25.39.$$

这个数值等于下面三个 BMI 观测值的中位数：

$$65/1.6^2,60/1.5^2,70/1.7^2.$$

最优值是

$$\min=|65-x_1 1.6^2|+|60-x_1 1.5^2|+|70-x_1 1.7^2|=6.25.$$

为了与模型 $w=b$ 的最佳 l^1 拟合进行比较，先计算中位数 $b=x_1=65$，对应的最优值是

$$\min=|65-x_1|+|60-x_1|+|70-x_1|=10.$$

所以，对于题目中所给的数据，根据 l^1 拟合的结果，模型 $w=ch^2$ 优于模型 $w=b$.

$p=2$ 的情形：这个优化问题可以转化为求解关于 c 的线性方程(见下面的第 23 节). 在此我们用微积分知识求解，需要用到下面的目标函数可微这样一个事实：

$$f(c)=(65-1.6^2c)^2+(60-1.5^2c)^2+(70-1.7^2c)^2.$$

令 $f'(c)=0$，方程两边同时除以 -2，得到

$$1.6^2(65-1.6^2c)+1.5^2(60-1.5^2c)+1.7^2(70-1.7^2c)=0,$$

所以，最优解是 $c=x_2=2\,518\,500/99\,841\approx 25.225$.
这个数值 x_2 不等于三个 BMI 观测值的均值(均值大约是 25.426)，最优值大约等于 18.

体重 w_i 的均值是 65，对应的最小值是 $5^2+0^2+5^2=50$. 所以，模型 $w=ch^2$ 仍然优于模型 $w=b$.

$p=\infty$ 的情形：这个优化问题可以转化为含有两个变量的线性规划问题，然后用图解法或单纯形法求解(见下面的第 23 节). 我们也可以用只有一个变量的图解法来求解，需要极小化的目标函数是：

236

$$f(c)=\max(|65-1.6^2c|,|60-1.5^2c|,|70-1.7^2c|).$$

函数 $f(c)$ 是一个分段仿射凸函数，其函数图形如图 22.8 所示 .

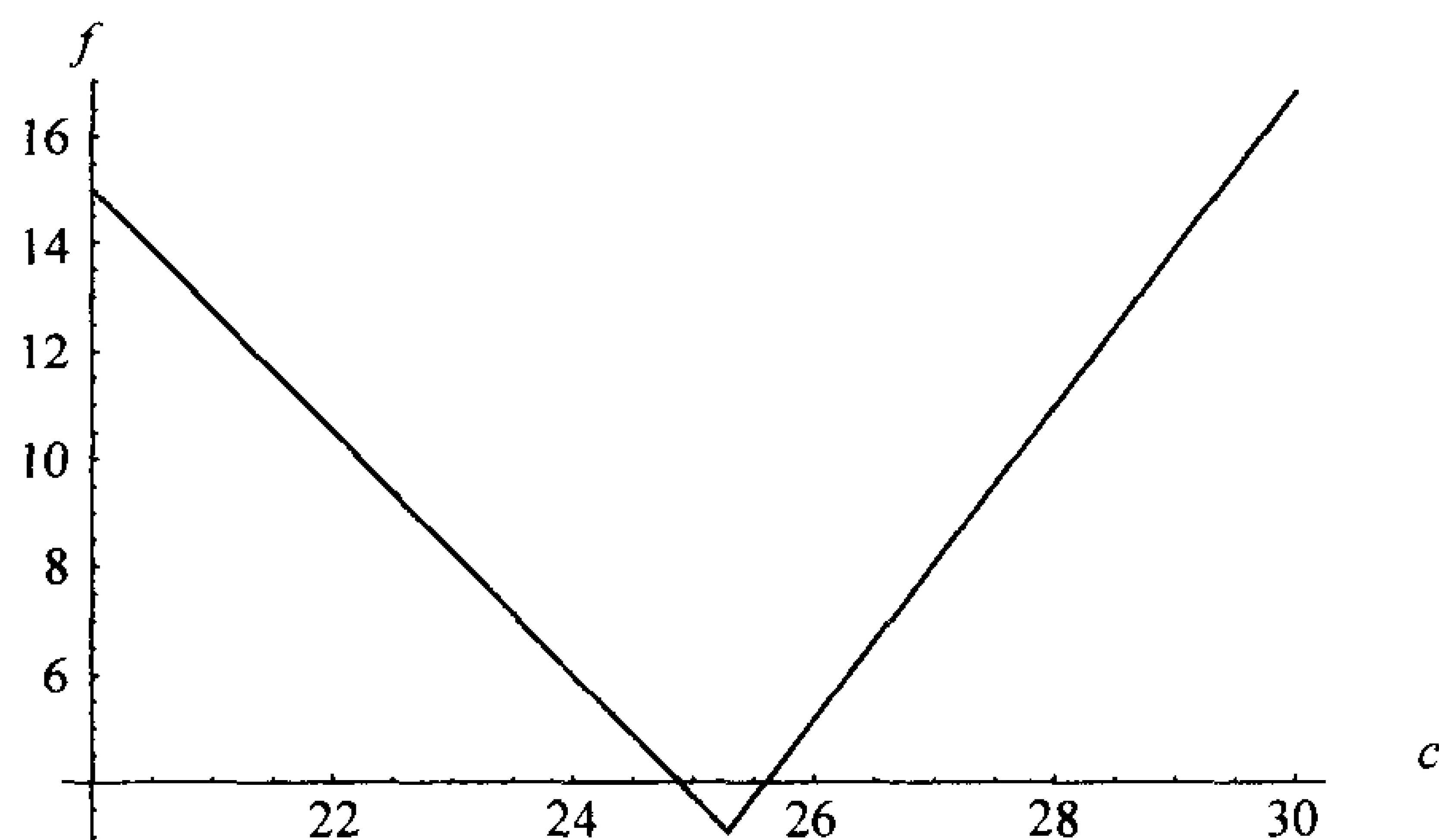


图 22.8 l^∞ 拟合的目标函数

从图中可以看出最优解 $c\approx25$. 在这一点，

$$65-1.6^2c\approx1, 60-1.5^2c\approx3.75, 70-1.7^2c\approx-2.25;$$

因此

$$f(c)=\max(60-1.5^2c, -70+1.7^2c).$$

所以，精确的最优解应该满足 $60-1.5^2c=-70+1.7^2c$，即最优解是

$$c_\infty=6\,500/257\approx25.29.$$

这个数值不等于三个 BMI 观测值的中程数(中程数大约是 25.44)，最优值大约等于 3.

另一方面，体重 w_i 的中程数是 65，即模型 $w=b$ 的最佳 l^∞ 拟合对应的最小值是 5. 所以，模型 $w=ch^2$ 仍然优于模型 $w=b$. ■

注 22.9 在上述例子中，对于 $p=1, 2, \infty$，最优解各不相同. 对于不同的 p，比较对应的最优值 M_p 没有意义，除非先对它们进行标准化：

$$\text{当 } p\neq\infty \text{ 时 } M_p\mapsto(M_p/m)^{1/p}, M_\infty\mapsto M_\infty,$$

其中 m 是观测值的数量(问题 22.7 中 $m=3$). 这个变换将残差列向量的 l^p 范数转换成残差的绝对值的平均值. 但是，即使经过了这样的变换，对不同的 p，也很难说清比较这些平均数的道理所在.

237

线性逼近问题的一般提法

一般地，给定了以下数据：m 个元素组成的一个列向量和一个 $m\times n$ 的矩阵. 希望找到 n 个元素组成的一个列向量 X，使得残差列向量 $e=(e_i)=w-AX$ 在下面的范数意义下是最小的：

$$\|e\|_2=(\sum e_i^2)^{1/2}, \quad \|e\|_1=\sum|e_i|, \quad \|e\|_\infty=\max(|e_i|).$$

这样，实际上我们定义了三个优化问题，其中有 m 个变量，且目标函数是非线性的. 这类问题在统计学中非常典型，也经常在数学的其他领域和计算机科学中出现.

注 22.10 从几何上看，这个问题实际上是在矩阵 A 的列空间中找一个向量

来逼近向量 w(根据定义，列空间是指 A 的列向量的所有线性组合组成的集合). 在下一节中，我们可以看到，第一个极小化问题实际上是求解一个线性方程组，而另外两个问题可以归结为线性规划问题.

在前面介绍的定理和例子中，我们考虑过矩阵 A 只含有一列的情况(即 $n=1$). 事实上，在前面的定理中，矩阵 A 可以认为是由 m 个 1 组成的一个列向量，并且这种情况下的问题根据定理已经被完全解决. 当 $n=1$ 而 A 是一般的列向量时，寻找最优拟合的问题是一个无约束的单变量优化问题. 正如问题 22.7 所示，如果不用计算机，求解这样的问题可能具有一定的挑战性.

前面提到的例子 $w=a+bh+ch^2+dh^3$(残差是 $e_i=w_i-(a+bh_i+ch_i^2+dh_i^3)$)是符合线性逼近一般提法的，即 $n=4$，残差是未知数 X 的线性函数. 科学家用于数据拟合的函数可以是他们所知道的任何函数(如多项式函数、有理函数、三角函数、指数函数等)，但为了得到系数，我们或者只需要求解线性方程组(对于最小二乘逼近)，或者只需要使用线性规划就已足够.

注 22.11　与统计学的联系：线性逼近方法，特别是最小二乘法，在统计学中也会出现. 在传统的回归分析中，矩阵 A 的第一列是由 m 个 1 组成的一个列向量. 在简单的回归分析中，矩阵 A 含有 $n=2$ 个列. 在多重回归分析中，$n\geqslant 3$. 在时间序列分析中，A 的第二列是表示时间的算术级数，典型的情况是 $[1, 2, \cdots, m]^T$.

238

比较高深的概率工具基本上是在残差服从正态分布的假定下使用，此时残差被看作是具有某种对称特性的随机噪声. 这些假定可以说明最小二乘方法(l^2 拟合)的合理性，这是统计学中所采用的方法. 当 $n=1$ 时，这些假定倾向于在所有平均数中使用均值，不过样本中位数也被广泛使用.

虽然不能明确检验多重回归方法的假定条件，但对于严重违反这些假定的情况应进行相应的处理. 特别是异常点(也就是极端情况)的存在会使结果产生严重的偏差，即在特定的方向上“推拉”回归直线，导致回归系数的偏差. 经常遇到的情况是，即使只排除掉一个极端情况的数据点，也会产生完全不同的结果. 多重回归(这个术语由 Pearson 于 1908 年最早使用)的一般目的是希望得到几个独立变量或预测变量与一个相关变量或标准变量之间更多的关系. 从统计学的观点看，为了可靠地估计列向量 X 的 n 个未知系数，观测值的个数 m 应该远远大于变量的个数 n.

练习

1～4. 计算下面给出的数的均值、中位数和中程数.

1. 2，－7，0，2，1.

2. 23，56，－6，0，8，0，67.

3. 2，－7，0，2，1，0，0，－1，8.

4. 2，3，5，6，－6，0，8，0，6，7.

5. 构造满足下面条件的数值例子，其中 x_2 是均值，x_1 是中位数，x_∞ 是中程数.

(a) $x_2<x_1<x_\infty$，　(b) $x_1<x_2<x_\infty$，

(c) $x_2<x_\infty<x_1$，　(d) $x_\infty<x_2<x_1$，

(e) $x_\infty<x_1<x_2$，　(f) $x_1<x_\infty<x_2$.

6. 注意在问题 22.7 中，通过简单地计算 BMI 的中位数就可以得到结果 $a_1=65/1.6^2\approx25.39$. 这是一定成立的(这会使得所有形如 $w=ax$ 的模型的最佳 l^1 拟合的计算变得非常简单)还是一个偶然的巧合？提示：试试其他的例子. 由于目标函数是分段线性的、非常数的和非负的，因此为了找到最优点，我们只需要考虑斜率发生变化的点，即给定的 BMI 的点 w_i/h_i^2.

7. 利用问题 22.7 中的数据，对 $p=1,2,\infty$ 和模型 $w=ah$，计算最佳的 l^p 拟合. 与问题 22.7 中的最佳拟合结果进行比较. 239

8. 利用问题 22.7 中的数据，对 $p=1,2,\infty$ 和模型 $w=ah^3$，计算最佳的 l^p 拟合. 与问题 22.7 中的最佳拟合结果进行比较.

注　这个模型是文献中提出的. w/h^3 这个量被称为标准体重(normalized body mass，NBM)，被建议用来代替 BMI(w/h^2). BMI 是由比利时统计学家和人类学家 Lambert-Adolphe-Jacques Quetelet (1796—1874)提出的. BMI 和 NBM 中使用同样一些度量单位. 文献中也提出了形如 $h/w^{1/3}=\text{NBM}^{-1/3}$ 的重量指数(ponderal index)，还有建议使用 w/h^4.

9. 利用下面的数据，对 $p=1,2,\infty$ 和模型 $w=ah^2$，计算最佳的 l^p 拟合：

i	1	2	3	4
身高 h(m)	1.6	1.5	1.7	1.8
体重 w(kg)	65	60	70	80

与模型 $w=b$ 的最佳拟合(用相同的 p)的结果进行比较.

10. 利用第 9 题中的数据，对 $p=1,2,\infty$ 和模型 $w=ah$，计算最佳的 l^p 拟合.

11. 利用第 9 题中的数据，对 $p=1,2,\infty$ 和模型 $w=ah^3$，计算最佳的 l^p 拟合.

12. 下面是前 100 个质数 p_n：

2，3，5，7，11，13，17，19，23，29，31，37，41，43，47，53，59，61，67，71，73，79，83，89，97，101，103，107，109，113，127，131，137，139，149，151，157，163，167，173，179，181，191，193，197，199，211，223，227，229，233，239，241，251，257，263，269，271，277，281，

283，293，307，311，313，317，331，337，347，349，353，359，367，373，379，383，389，397，401，409，419，421，431，433，439，443，449，457，461，463，467，479，487，491，499，503，509，521，523，541.

对 $p=1$，2，∞和模型 $p_n=cn$，计算最佳的 l^p 拟合.

13. 利用第 12 题中的数据，对 $p=1$，2，∞和模型 $p_n=cn\log(n)$，计算最佳的 l^p 拟合.

14. 斐波那契数 F_t 是由下面的递归式定义的：

$$F_t = F_{t-1} + F_{t-2}, F_0 = F_1 = 1.$$

用前 50 个斐波那契数，对 $p=1$，2，∞，按照模型 $F_t=2^{ct}$ 计算最佳的 l^p 拟合.

240

23. 线性逼近与线性规划

众所周知，最佳 l^2 拟合就是最小二乘拟合. 它之所以被广泛使用，主要有以下两个原因：首先，在残差服从一定概率分布的假定下，可以证明其合理性；其次，它的计算相对比较容易. 现在我们回忆一下如何通过求解线性方程组进行求解(这可以看成是线性规划的一种特殊形式).

最佳 l^2 拟合

欧氏范数 $\|e\|_2=\sum e_i^2$ 是最常用于衡量一个向量大小的范数，而且在欧氏几何中使用的就是这个范数. 为了在列空间中找到一个向量，我们从 w 向列空间作一条垂线(见注 22.10). 换言之，我们希望向量 $w-AX$ 与 A 的所有列正交，即 $A^T(w-AX)=0$. 这就给出了一个关于列向量 X 的含有 n 个未知数的 n 个线性方程 $A^TAX=A^Tw$，该方程组总是有解的. 此外，对于所有的解 X，最佳拟合 AX 总是一样的. 当 w 属于列空间时，最佳拟合就是 w. 否则解 X 是唯一的.

例 23.1 根据前面的一般讨论，假设 $n=1$，且列向量 A 的所有元素都是 1，那么，我们需要找到一个数 $X=a$，使得$\sum(w_i-a)^2$ 最小. 换言之，我们希望用一个数 a 近似 m 个给定数 w_i. 此时，方程 $A^TAX=A^Tw$ 成为$na=\sum w_i$，因此 $a=X=\sum w_i/n$ 是所给数的算术平均值. 这和定理 22.1 一致. ■

问题 23.2 对下面的数据，按照模型 $w=a+bh$，计算最佳的 l^2 拟合(保留两位小数).

i	1	2	3	4	5
h	1.5	1.6	1.7	1.7	1.8
w	60	65	70	75	80

解 根据前面的一般讨论，这里 $X=[a, b]^T$，$w=[60, 65, 70, 75, 80]$，

$$A^T=\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 1.5 & 1.6 & 1.7 & 1.7 & 1.8\end{bmatrix},$$

方程组 $A^TAX=A^Tw$ 的形式是

$$\begin{bmatrix}5 & 8.3\\ 8.3 & 13.83\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}=\begin{bmatrix}350\\ 584.5\end{bmatrix}.$$

241

求解这个方程组得到 $a\approx-41.7$，$b\approx67.3$. ■

问题 23.3 对问题 23.2 中的数据，按照模型 $w=a+ch^2$，计算最佳的 l^2 拟合(保留两位小数).

解 根据前面的一般讨论，这里 $X=[a, c]^T$，$w=[60, 65, 70, 75, 80]$，

$$A^T=\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 1.5^2 & 1.6^2 & 1.7^2 & 1.7^2 & 1.8^2\end{bmatrix},$$

方程组 $A^TAX=A^Tw$ 的形式是

$$\begin{bmatrix}5 & 13.83\\ 13.83 & 38.82\end{bmatrix}\begin{bmatrix}a\\ c\end{bmatrix}\approx\begin{bmatrix}350\\ 979.65\end{bmatrix}.$$

求解这个方程组得到 $a\approx13.37$，$c\approx20.47$. ■

最佳 l^1 拟合

此时优化问题是 $\|e\|_1\to\min$，其中 $e=(e_i)=w-Aa$. 通过引入 m 个新的变量 u_i，增加约束 $|e_i|\leqslant u_i$(对所有的 i)，这个优化问题可以归结为线性规划. 于是，可以得到如下线性规划，它有 $m+n$ 个变量 a_j，u_i，有 $2m$ 个线性约束：

$$\sum u_i\to\min,\ -u_i\leqslant w_i-A_ia\leqslant u_i,\quad i=1, \cdots, m,$$

其中 A_i 是 A 的第 i 行.

问题 23.4 对问题 23.2 中的数据，按照模型 $w=bh$，计算最佳的 l^1 拟合.

解 这里 $A=[1.5, 1.6, 1.7, 1.7, 1.8]^T$. 如前所述，可以把这个问题转化成 6 个变量的线性规划问题，然后用单纯形法求解. 不过，我们也可以直接考虑非线性规划问题，即让目标函数 $f=\sum|w_i-bh_i|$ 最小，这时只含有一个变量 b，没有约束条件，因此可以用如下的图解法求解. 首先计算 $c_i=w_i/h_i$，得到

$$c_1=40, c_2=65/1.6\approx40.6, c_3=70/1.7\approx41.2,$$

$$c_4=75/1.7\approx44.1, c_5=80/1.8\approx44.4.$$

242

如果 $b\leqslant c_1=40=\min(c_i)$，那么 $f=\sum(w_i-bh_i)$，其斜率是

$$-\sum h_i=-h_1-h_2-h_3-h_4-h_5=-8.3.$$

在下一个区间 $c_1\leqslant b\leqslant c_2$ 上，f 的斜率是

$$h_1-h_2-h_3-h_4-h_5=-5.3.$$

在下一个区间 $c_2 \leqslant b \leqslant c_3$ 上，f 的斜率是

$$h_1 + h_2 - h_3 - h_4 - h_5 = -2.1.$$

在下一个区间 $c_3 \leqslant b \leqslant c_4$ 上，f 的斜率是

$$h_1 + h_2 + h_3 - h_4 - h_5 = 1.3.$$

对更大的 b，斜率也更大．所以容易看到，斜率在 $b=c_3$ 处改变符号．因此最小值在 $b=c_3=70/1.7 \approx 41.2$ 取到，所以答案为 $w=70h/1.7$. ■

注 最早出现的 l^1 逼近问题与星体运动的数据有关．Bosovitch（大约在1756年），Laplace(1789)，Gauss(1809)，Fourier(大约在1822年)提出了求解这些问题的方法．实际上，Fourier也考虑了 l^∞ 逼近问题，并提出了找到任意一个线性约束系统可行解的方法．奇怪的是，虽然发现最佳 l^2 逼近问题能够归结为求解线性方程组，但是关于 l^2 逼近的研究工作却直到19世纪才出现(Legendre，Gauss).

最佳 l^∞ 拟合也称为最小绝对偏差拟合或切比雪夫逼近.

最佳 l^∞ 拟合

此时优化问题是 $\|e\|_\infty \to \min$，其中 $e=(e_i)=w-Aa$．通过引入一个新的变量 u，增加约束 $|e_i| \leqslant u$(对所有的 i)，这个优化问题可以归结为线性规划．类似的技巧在矩阵对策归结为线性规划时也用到过．于是，可以得到如下线性规划，它有 $n+1$ 个变量 a_j，u，有 $2m$ 个线性约束：

$$t \to \min, -u \leqslant w_i - A_i a \leqslant u \text{ 对 } i = 1,\cdots,m,$$

其中，A_i 是 A 的第 i 行.

问题 23.5 对问题23.2中的数据，按照模型 $w=ah$，计算最佳的 l^∞ 拟合.

解 可以将这个问题转化成2个变量 a，u 与10个约束的线性规划问题，然后用图解法或单纯形法求解．或者也可以直接作出目标函数 $f=\max(w_i-ah_i)$ 的图形(图23.6).

243

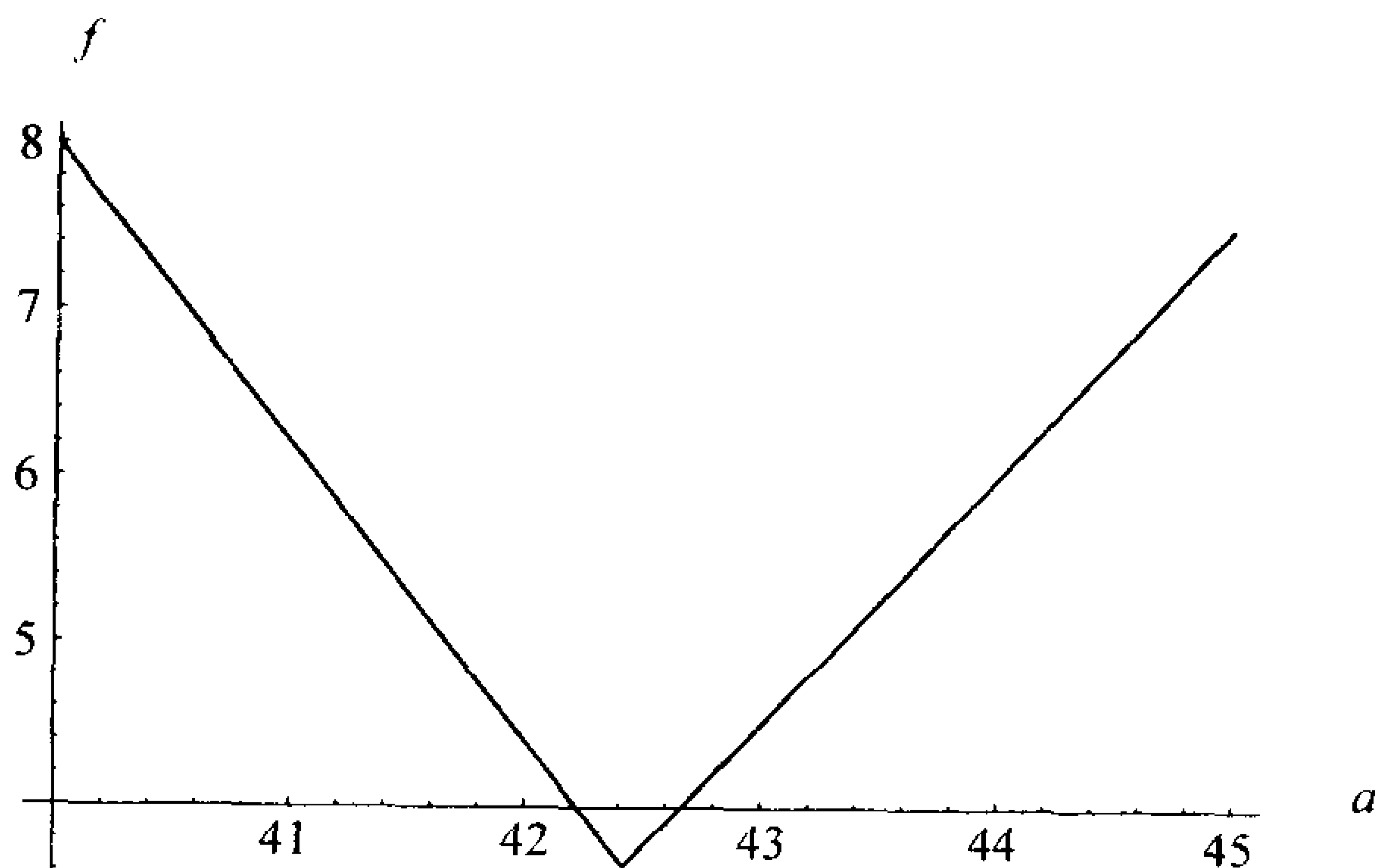


图 23.6 $f=\max(w_i-ah_i)$

可以看出最优解是$a\approx 42.4$. 为了得到更精确的答案，我们可以在$a=42.4$附近作出所有五个函数(w_i-ah_i)的图形，f应该是这五个函数中最大者(图 23.7).

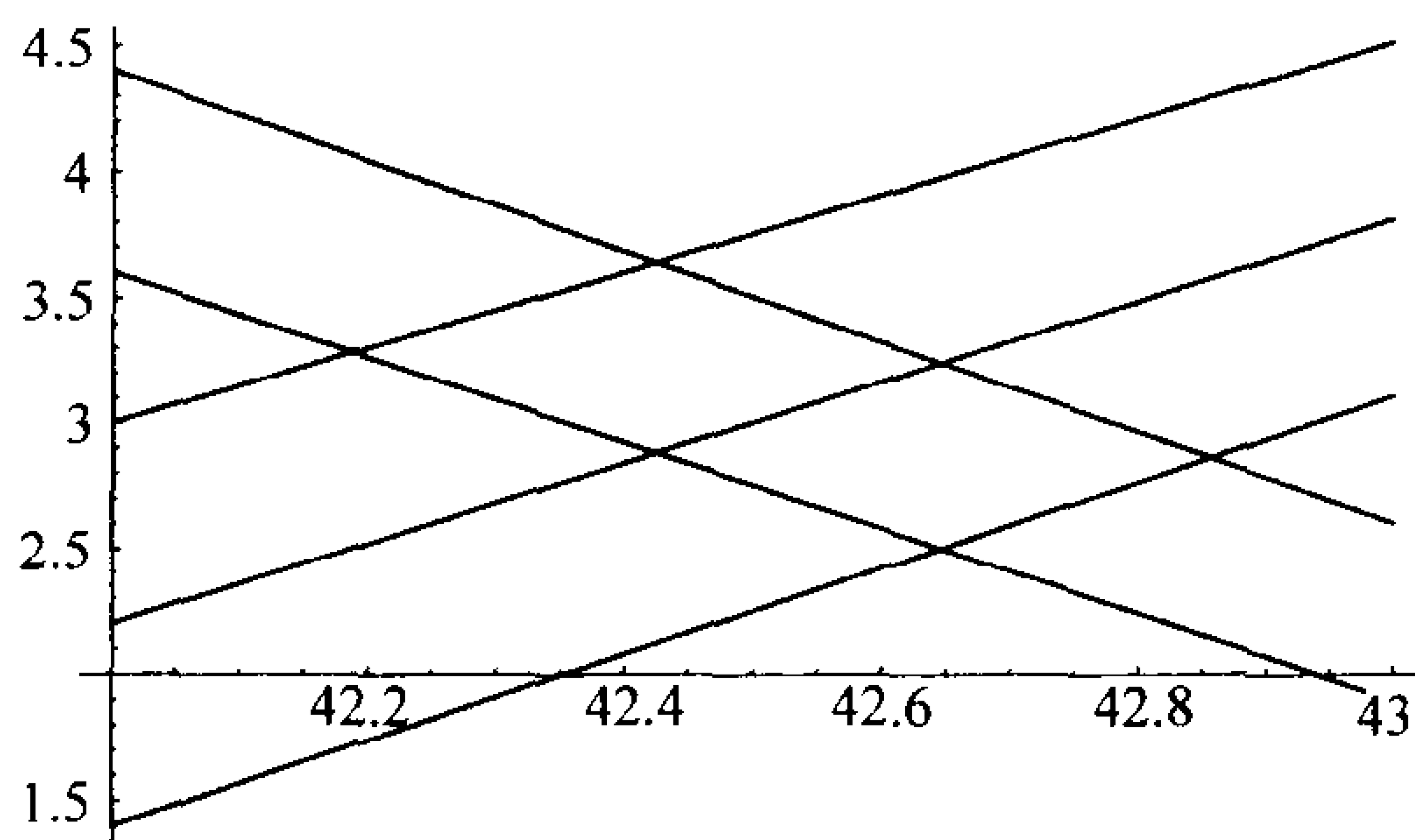


图 23.7

可以看出，在$a=42.4$附近，目标函数f是

$$\max(1.5a-60,80-1.8a),$$

所以最优解是

$$a=140/3.3\approx 42.42.$$

■

问题 23.8 利用下面给出的每小时测量得到的放射性水平数据，计算放射性同位素的半衰期：

小时	t	0	1	2	3	4
放射水平	r	100	76	58	45	34

244

解 本题所要求的是半衰期λ，显然，不用计算就可以知道$2<\lambda<3$. 在t小时后的放射性水平应该是$c2^{-t/\lambda}$，在排除测量误差和舍入误差以后，所给的数应该形成一个几何级数. 但是，模型$r=c2^{-t/\lambda}$关于λ不是线性的. 两边取对数，可以变成如下线性关系，其中$w=\log_2 r$，$a=\log_2 c$，$b=1/\lambda$：

$$w=a-tb. \tag{23.9}$$

有几种方法可以继续做下去，其中最简单的方法是计算r_t/r_{t-1}的一个平均数. 由此得到2^b的如下估计：均值 1.309 64，中程数 1.300 79，中值 1.319 66，对应的λ估计如下：2.569 58，2.635 8，2.498 96.

现在我们用线性模型(23.9)求解这个问题. 最佳的l^2拟合可以通过求解线性方程组$A^{\mathrm{T}}AX=A^{\mathrm{T}}w$得到，其中：

$$w=[\log_2 100,\log_2 76,\log_2 58,\log_2 45,\log_2 34]^{\mathrm{T}}, \tag{23.10}$$

$$X=\begin{bmatrix} a \\ -b \end{bmatrix},\quad A^{\mathrm{T}}=\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}.$$

所以方程组是

$$\begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}\begin{bmatrix} a \\ -b \end{bmatrix}=\begin{bmatrix} \log_2 674\,424\,000 \\ \log_2 31\,133\,130\,272\,352\,000 \end{bmatrix}\approx\begin{bmatrix} 29.329\,1 \\ 54.789\,3 \end{bmatrix}.$$

求解这个方程组，得到 $b\approx 0.387$；$\lambda\approx 2.58$.

现在求模型(23.9)的最佳 l^∞ 拟合. 对应的线性规划是

$$u\to\min,\ -u\leqslant w_t-a+bt\leqslant u\ 对\ t=0,1,2,3,4,$$

其中 w_t 是(23.10)中列向量 w 的元素. 这个问题有三个未知数 a，b，u，有十个约束. 求解得到 b 的最优值是 $b\approx 0.385\,259$，所以 $\lambda\approx 2.596$.

下面求模型(23.9)的最佳 l^1 拟合. 对应的线性规划是

$$u_0+u_1+u_2+u_3+u_4\to\min,$$

245

$$-u_t\leqslant w_t-a+bt\leqslant u_t,对\ t=0,1,2,3,4,$$

其中 w_t 是(23.10)中列向量 w 的元素. 这个问题有七个未知数 a，b，u_0，u_1，u_2，u_3，u_4，有十个约束. 求解该问题得到 b 的最优值是 $b\approx 0.389\,098$，所以 $\lambda\approx 2.570$.

除非我们知道产生这些数据的更多信息，否则很难确定哪个模型更好，不过 $\lambda=2.55\pm 0.05$ 似乎是合理的. ■

问题 23.11 用三个样本对一个放射性计数器进行校准，三个样本分别是 1mg 的同位素 A，1mg 的同位素 B，1mg 的同位素 C：

小时	t	0	1	2	3	4
1mg 同位素	A	4 100	510	64	8	1
1mg 同位素	B	1 300	320	80	20	5
1mg 同位素	C	160	80	40	20	10

然后，用这个计数器测量其他样本的同位素含量. 假设某个样本 S 由同位素 A，B，C 和某些非放射性元素组成，下面是计数器的计数结果：

小时	t	0	1	2	3	4
放射水平	r	100	76	58	45	34

找出样本中同位素 A，B，C 的含量(mg).

解 设 a，b，c 为样本中同位素 A，B，C 的含量. 除非我们知道同位素之间发生核反应的一些信息，否则我们只好假定放射性水平是：

$$100=4\,100a+1\,300b+160c+e_0,对\ t=0,$$

$$76=510a+320b+80c+e_1,对\ t=1,$$

$$58=64a+80b+40c+e_2,对\ t=2,$$

$$45=8a+20b+20c+e_3,对\ t=3,$$

$$34=a+5b+10c+e_4,对\ t=4,$$

其中 e_0，e_1，e_2，e_3，e_4 为小误差.

最佳的 l^p 拟合是：

当 $p=2$ 时，$a\approx0.12$，$b\approx-0.65$，$c\approx2.74$；

当 $p=1$ 时，$a\approx0.13$，$b\approx-0.70$，$c\approx2.89$；

当 $p=\infty$时，$a\approx0.16$，$b\approx-0.78$，$c\approx2.96$. ■

我们看到，答案根据目标的不同而不同. 但在所有三种情形下，b 的值都是负数，这是无法接受的. 我们需要对未知数增加符号限制，以防止这种现象发生. 对于 l^1 拟合和 l 拟合，此时问题仍然是线性规划，但这个额外的符号约束将会使 l^2 拟合从线性规划问题变为非线性规划问题. [246]

但是 b 的负值并非很接近于 0，这是一个很强的信号，可能我们的解法或者我们的原始数据或者我们的假定是错的. 说到假定，在样本中是否可能还存在一种我们不知道的同位素呢？

关于解法，如果是采用计算机求解，我们是否按照软件的要求正确地输入了数据？这个软件是否有漏洞，所以得到的解是错误的？计算机是否出现了随机的错误？

下面我们使用 Mathematica 软件，介绍两种不同的方法求解最佳 l^2 拟合. 第一种方法是使用命令“FindMinimum”. 首先导入数据(如 e0 $=e_0$，等)：

```
e0 =- 100 + 4 100a + 1 300b + 160c;
e1 =- 76 + 510a + 320b + 80c;
e2 =- 58 + 64a + 80b + 40c;
e3 =- 45 + 8a + 20b + 20c;
e4 =- 34 + a + 5b + 10c;
```

(分号表示不要求软件打印录入的数据，这时我们可以在一个输入块中输入多个数据.)然后键入：

```
FindMinimum[e0^2 + e1^2 + e2^2 + e3^2 + e4^2],
    {a,1},{b,1},{c,1}
```

得到的响应如下：

```
{158.785,{a -> 0.124 638,b ->- 0.653 399,c -> 2.741 08}}
```

输出中的第一个数 158.785 是最优值(±0.005)；{a，1}，{b，1}，{c，1}中的“1”表示搜索最优解时迭代法采用的初始点.

第二种方法是通过求解线性方程组 $A^{T}AX=A'B$ 得到最小二乘拟合，其中 $X=[a, b, c]^{T}$，

$$[A/B]=\begin{bmatrix} 4\,100 & 1\,300 & 160 & | & 100 \\ 510 & 320 & 80 & | & 76 \\ 64 & 80 & 160 & | & 58 \\ 8 & 20 & 20 & | & 45 \\ 1 & 5 & 10 & | & 34 \end{bmatrix}.$$

将 A，X，B 如下输入 Mathematica：

X={a, b, c}; B={100, 76, 58, 45, 34}; A={{4 100, 1 300, 160}, {510, 320, 80}, {64, 80, 40}, {8, 20, 20}, {1, 5, 10}};

在 Mathematica 中，方程组 $A^{\mathrm{T}}AX=A'B$ 是

[247]

Transpose[A]. A. X == Transpose[A]. B

此时可以用“Solve”命令进行求解：

Solve[Transpose[A]. A. X == Transpose[A]. B,{a,b,c}]

输出结果为：

$$\left(\left\{\left\{a->\frac{69\,369\,996}{556\,572\,001}, b->-\frac{10\,909\,908\,649}{16\,697\,160\,030}, c->\frac{91\,536\,610\,621}{33\,394\,320\,060}\right\}\right)$$

为了得到十进制数的显示结果，输入命令“N[%]”得到

{{a --> 0. 124 638,b -->- 0. 653 399,c --> 2. 741 08}}.

这与使用“FindMinimum”得到的结果相同. 注意，Mathematica 中也有其他命令可以计算最小二乘拟合.

为了进一步检查这个结果是否正确，下面用 Maple 软件求解这个问题. 将矩阵输入 Maple 的方式是：

```
A: =array([[4 100,1 300,160],[510,320,80],
     [64,80,40],[8,20,20],[1,5,10]]);
B: =array([100,76,58,45,34]);
```

计算最小二乘拟合的命令是

```
with(linalg);
leastsqrs(A,B);.
```

输出结果是：

```
[69369996    -10909908649   91536610621]
[---------,  ------------,  -----------]
[556572001    16697160030   33394320060]
```

这与用 Mathematica 得到的结果相同. 所以，可能我们的原始数据或者我们的假

定错误.

问题 23.12 下面是牛津剑桥大学 10 名学生的 SAT(Scholastic Aptitude Test，学术能力测验)分数和 GPA(Grade Point Average，平均积分点)的数据：

SAT1	x	750	720	710	780	700	730	760	770	720	720
SAT2	y	740	730	710	770	720	740	770	760	710	730
GPA	z	3.5	3.4	3.6	3.7	3.2	3.2	3.8	3.7	3.5	3.4.

对 $p=1,2,\infty$ 和模型 $z=ax+by$，计算最佳的 l^p 拟合：

解 令

$$A=\begin{bmatrix}750 & 720 & 710 & 780 & 700 & 730 & 760 & 770 & 720 & 720\\ 740 & 730 & 710 & 770 & 720 & 740 & 770 & 760 & 710 & 730\end{bmatrix}^{\mathrm{T}},$$

248

$$b=[3.5,3.4,3.6,3.7,3.2,3.2,3.8,3.7,3.5,3.4]^{\mathrm{T}},$$

$$e=A\begin{bmatrix}x\\ y\end{bmatrix}-b.$$

我们的三个优化问题是

$$\|e\|_p\rightarrow\min,\text{对 } p=1,2,\infty.$$

三个最优解是：

$$x\approx 0.008,y\approx -0.004,\text{对 } p=2;$$

$$x\approx 0.015,y\approx -0.010,\text{对 } p=\infty;$$

$$x\approx 0.008,y\approx -0.003,\text{对 } p=1.$$ ■

这里的负值不像上一个例子离 0 点那么远，似乎并非不可想象. 但这个负值的含义是，第二次 SAT 获得的一个较高的成绩反而会使 GPA 减少，这个结论非常值得怀疑. 所以说，或者我们的模型选得不合适，或者所给的数据不够，不能得到任何有用的结论. 有无数模型和计算结果最后直接进了垃圾桶，没有发表出来. 即使发表出来的模型和计算结果，有些也属于垃圾一类.

一般来说，对计算结果的解释是应用数学非常重要的一个组成部分. 关于收集和处理数据、用数据进行计算、对计算结果进行解释等，更多的细节可以在统计学的教科书中找到.

练习

1. 对问题 22.7 中的数据，找到最小二乘拟合模型 $w=a+bh$. 将这个拟合的最优值与模型 $w=ah^2$ 的最小二乘拟合的最优值进行比较.

2. 求出例 6.10 中线性系统的最小二乘解. 回忆一下，这个系统无解. 一般来说，方程组 $Ax=b$ 的最小二乘解 $\hat{x}$ 并不是解(即一般来说 $A\hat{x}\neq b$). 不过，$A\hat{x}$ 将尽可能地靠近 b.

3. 求出下面线性系统的最小二乘解：

$$\begin{array}{cc} a & b \end{array}$$
$$\begin{bmatrix} 3 & 4 \\ -1 & 5 \\ 3 & 0 \\ 1 & -7 \end{bmatrix} \begin{array}{l} =1 \\ =2 \\ =4 \\ =5 \end{array}$$

249

4. 对第 22 节练习 9 的数据，找到最小二乘拟合模型 $w=a+bh^2$.

5. 对问题 23.2 中的数据，找到形为 $w=ah^2$ 的最佳 l^1 拟合.

6. 对问题 23.2 中的数据，找到形为 $w=ah^3$ 的最佳 l^1 拟合.

7. 对问题 23.2 中的数据，找到形为 $w=ah+b$ 的最佳 l^1 拟合.

8. 将问题 23.2 中的最小值与第 5、6 题的最小值进行比较，并指出对于所给的数据，哪个模型给出的拟合效果最好.

9. 对问题 23.2 中的数据，找到形为 $w=ah^2$ 的最佳 l^∞ 拟合.

10. 对问题 23.2 中的数据，找到形为 $w=ah^3$ 的最佳 l^∞ 拟合.

11. 对问题 23.2 中的数据，找到形为 $w=ah+b$ 的最佳 l^∞ 拟合.

12. 对第 22 节练习第 12 题的数据，找到形为 $p_n=an+b$ 的最佳 l^2 拟合，并将结果与拟合 $p_n=n\log(n)$进行比较，其中 log 表示的是自然对数(这个拟合没有参数).

注 已经知道，当 $n\to\infty$ 时 $p_n/(n\log(n))\to 1$.

13. 对第 22 节练习第 13 题的数据，找到形为 $F_n=a2^{bn}$ 的最佳 l^2 拟合，并将结果与拟合$F_n=\alpha^{n+1}/\sqrt{5}$进行比较，其中 $\alpha=(\sqrt{5}+1)/2$(这是黄金分割比).

注 已经知道，对所有 t 有 $F_t=(\alpha^{t+1}-(-1/\alpha^{t+1}))/\sqrt{5}$.

14. 把下面的问题改写成线性规划，并求解这个问题：

$$|e_1|+2|e_2|+\max[|e_3|,|e_4|]\to \min$$

subject to
$$e_1=2x_1+3x_2-1.$$
$$e_2=x_1-2x_2-2.$$
$$e_3=-x_1+x_2+3.$$
$$e_4=x_1-x_2-4.$$
$$e_1\leqslant|e_3|\leqslant 4.$$

250

24. 其他例子

在下面的例子中，我们将讨论如何选择合适的模型进行拟合，如何解释计算的结果，而不把注意力集中在对应的优化问题的求解上，因为正如在 23 节中解

释过的，求解的问题可以归结为线性规划．这里的数据都不是人为构造的，而是要么来自于网上，要么来自于数学方面的研究问题．

例 24.1（时间序列） 假设你对 1995 年日本人每人每年消费的巧克力量(g)感兴趣，但你只知道此前 10 年的数据：

1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
1253	1313	1394	1535	1590	1535	1648	1626	1585	1499

如何使用这些数据（在本书中显然不希望使用传统的方式，如占卜用的纸牌）预测 1995 年的消费量？能找到好的方式回答这类问题的人，在股票市场上就能赚到大钱．

我们尝试一下简单的模型 $w=ah+b$，其中 h 是年，而 w 是每人每年消费的巧克力量(g)．也许实际上 w 还依赖于气候、每人的年收入、巧克力的价格、健康方面的考虑等因素，但这里假设我们没有关于这些因素的任何数据．我们对 $p=1$，2，∞计算最佳的 l^p 拟合，而不考虑其他拟合．下面是我们用计算机求解得到的答案，其中 1995 年的预测是 $w_{11}=1995a+b$：

$$a \approx 46.4, b \approx -90\,802.8, w_{11} \approx 1\,765.2, \text{对 } p=1,$$
$$a \approx 33.7, b \approx -65\,548.4, w_{11} \approx 1\,683.2, \text{对 } p=2,$$
$$a \approx 27.333, b \approx -52\,888.2, w_{11} \approx 1\,641.83, \text{对 } p=\infty.$$

现在，读者自己可以判断这些拟合所预测的结果是否足够好(1995 年的实际值是 1566g)．本题的所有数据来自于网址 http：//202.167.121.158/ebooks/jetro/November.html#01.

图 24.2 给出了 $w-1253$ 与 $h-1984$ 的图形（包括 11 年的数据），同时给出了最佳的 l^∞ 拟合．

251

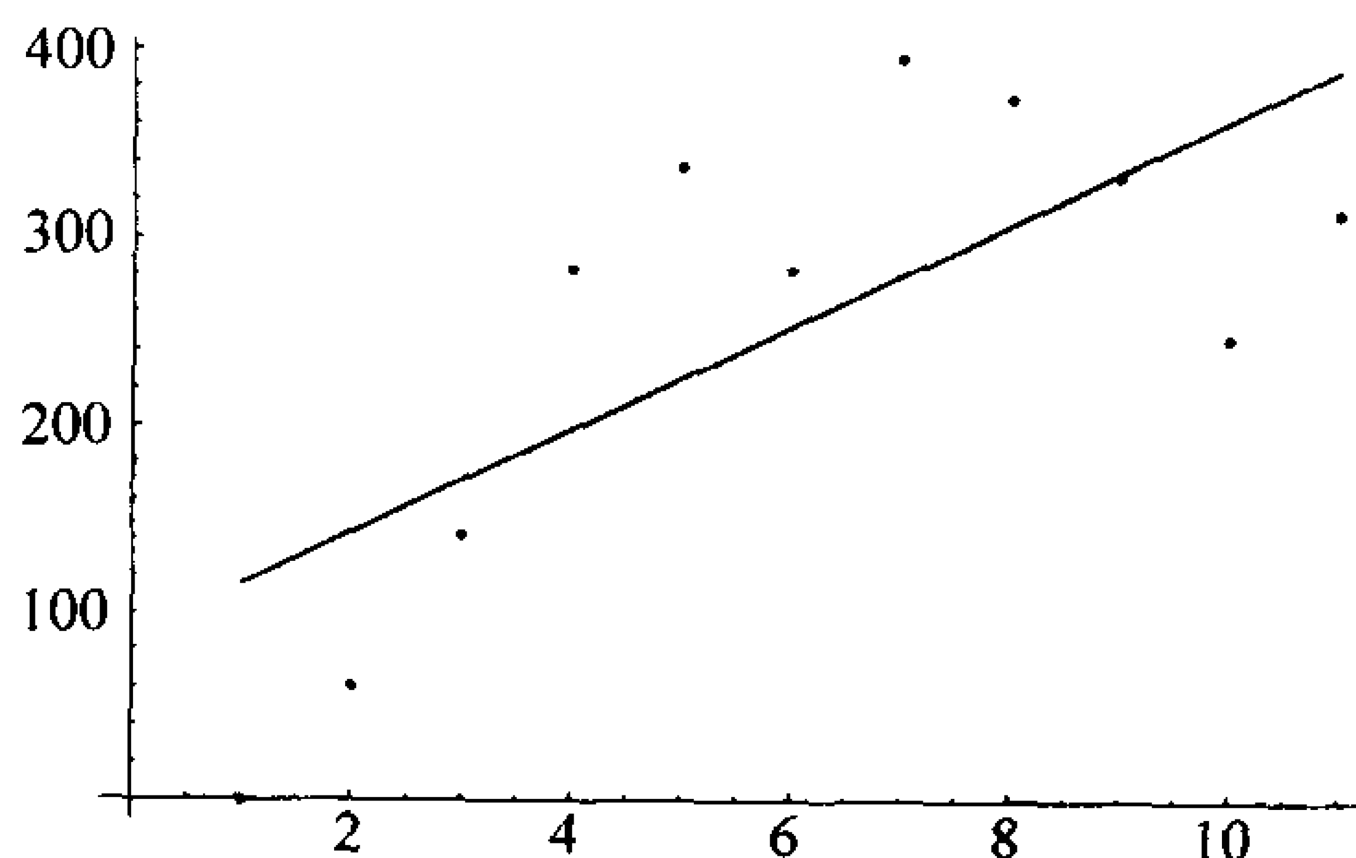


图 24.2 1985～1995 年日本人均每年巧克力的消费量 w(g，减去了 1253)

从图中可以看出（在进行计算之前先画图看看，通常是很有意义的），模型 $w=ah+b$ 的拟合效果并不好．似乎消费的趋势并不是按照常数比率 a 逐渐增加，相反增长量 w_i-w_{i-1} 在逐渐下降，到 1992 年变成了负值．所以趋势并不单调，一

个更好的模型可能是 $w_i - w_{i-1} = a'h + b'$，即 $w = ah^2 + bh + c$，其中 $a = a'/2$，$b = b' + a'/2$.

由于模型 $w = ah^2 + bh + c$ 比 $w = bh + c$ 更一般，所以对 1985～1994 年的拟合来说一定更好或至少一样好，但具体到对 1995 年的预测，却并不一定更好. 下面是计算结果：当 $p=1$ 时 $w_{11} \approx 1\,487.2$，当 $p=2$ 时 $w_{11} \approx 1\,453.1$，当 $p=\infty$ 时 $w_{11} \approx 1\,641.8$.

总是能够找到一个模型精确地预测你所要的答案. 然而我们希望，能够找到一个简单的模型，预测我们所不知道的答案. 对本题数据的解释参见网址：http：//202.167.121.158/ebooks/jetro/November.html#01 ■

例 24.3 有一位统计学家，对数字"1"在圆周率 $\pi = 3.141\,59\cdots$ 中出现的频率很感兴趣. 她计算了 π 的 20 亿位数字，记录了小数点后数字"1"出现的位置：

$$2, 4, 38, 41, 50, 69, 95, 96, 104, 111, 139, 149, 154, 155, 156, \cdots$$

她让我们用 ai 近似位置 w_i，其中未知数是 a. 她猜想 a 应该非常接近于 10.

针对所给出的 15 个数，我们能用学过的工具做点什么？首先，我们可以计算

252 $w_i/i(1 \leqslant i \leqslant 15)$ 的三个平均数.

均值是

$$2\,793\,047/270\,270 \approx 10.334\,3;$$

样本区间中值是

$$109/14 \approx 7.785\,71;$$

中值是

$$415/36 \approx 11.527\,8.$$

因此，我们对 $p=1, 2, \infty$，得到了形如 $w_i/i = a$ 的最佳 l^p 拟合，均值很靠近 10.

现在我们可以计算 $w_i = ai$ 的最佳 l^p 拟合，得到 $p=1$ 时 $a \approx 11.56$，$p=2$ 时 $a = 11.5$，$p=\infty$ 时 $a = 11$.

我们是否对 20 亿位数字尝试这些计算呢？那我们就只能再等若干年，等待功能更强大的计算机出现. ■

注 我们如何确认或拒绝这个猜想？注意，如果我们改变或丢掉一个无穷序列中的有限个数，序列的极限不会改变(如果极限存在的话).

例 24.4(单边拟合) 你把月工资(5 千美元支票，每个月的第一天拿到)以电子方式存入你的现金账户(money market account，MMA)，年利率为 3%. 利息是按照每月你的现金账户上的最小钱数计算的，并在 12 月 31 日进入你的现金账户. 在同一家银行你还有一个支票账户(checking account，CA)，这个账户没有利息. 你总是通过邮寄支票付账，出了支付邮票外从来不用现金，支付邮票的现

金从现金账户中提取. 所以你必须经常去银行，将现金账户中的钱转到支票账户，以便支付住房贷款、固定电话费、有线电视费、移动电话费、上网费、汽车保险、信用卡费、以及其他费用，每个月要使用 10 张支票，一共 3 千美元. 此外，每个季度(三月、六月、九月、十二月)你还要使用 5 张支票支付一些账单，一共 2 千美元；每年四月需要分别支付 2 千美元和 5 千美元的税单，每年八月还要支付 4 千美元的税单. 最后，在每年的12 月31 日，你还要签发本年度的最后 7 张支票，将你年收入中的其他部分(5 千美元减去邮资)捐给你最认可的基金会和慈善机构. 假设开始时你现金账户中的钱足够多，所以不需要担心超支.

现在银行提出让你改变这个程序：他们将从现金账户中支付所有的账单，没有手续费. 这样，你就不必每年填写 150 张支票和邮寄地址，可以节省邮资和支票. 此外，银行自动在每月的第一天将你现金账户上一定数量的钱转到支票账户上，自动在每年的第一天(1 月 1 日)将你现金账户一定数量的钱转到支票账户上，不需付出任何费用. 如果你接受这个方案，你就没有必要再去银行和邮局了. 但是，你必须决定每月转账和每年转账的数额 a 和 b. 253

如果你把每月的工资 5 千美元全部转到支票账户($a=5$, $b=0$)，一切都是可行的，但你不能从现金账户获得任何利息. 当 a 很小时，你的支票账户将会超支，可能引起罚款和其他一系列麻烦. 那么最优解是什么? 为此，我们把所有数据列入表 24.5.

表 24.5 例 24.4 中的数据和变量

月 t	支付 (千美元)	转账 (千美元)	余额 e_t(千美元)
1	3	$a+b$	$a+b-3$
2	3	a	$2a+b-6$
3	3+2	a	$3a+b-11$
4	3+7	a	$4a+b-21$
5	3	a	$5a+b-24$
6	3+2	a	$6a+b-29$
7	3	a	$7a+b-32$
8	3+4	a	$8a+b-39$
9	3+2	a	$9a+b-44$
10	3	a	$10a+b-47$
11	3	a	$11a+b-50$
12	3+2+5	a	$12a+b-60$

现在叙述你面对的优化问题，变量有两个：a，b，满足的约束是

$$a \geqslant 0, b \geqslant 0, 12a + b = 60.$$

此外，还有 12 个约束 $e_i \geqslant 0$(见表 24.5). 你不必加上保持现金账户平衡的约束，因为你的目标是从现金账户中得到的利息最大，这等价于使支票账户的余额 $e_1+\cdots+e_{12}$最小，因为支票账户没有利息. 254

我们将这个问题的求解过程留给读者(见练习第 3 题). 可以看到，这个问题和找极小化$\sum|e_i|$(其中 $e_i=ah_i+b-w_i$)的最佳 l^1 拟合是类似的. 但是，这里增加了约束 $e_i\geqslant 0$ 以及 a，$b\geqslant 0$. 这些符号限制可以帮助我们把优化问题改写成线性规划的规范形式，即 $\sum_i ah_i+b-w_i\rightarrow\min$,对所有 i 有 $ah_i+b-w_i\geqslant 0$.

■

练习

1. 下面是 9 年内美国新鲜草莓的产量 w(单位：百万磅). 分别对 $p=1$，2，∞，用模型 $w=ah+b$ 计算最佳 l^p 拟合，预测 1993 年的产量. 数据来自于 http://www.nalusda.gov/pgdic/Strawberry/ers/ers.htm.
将得到的预测值与实际值 987.6 比较. 提示：有些计算机软件不适合求解含有较大数据的问题，所以可以把 h 用 $h-1988$ 替换，产量 w 用 $w-x_2$ 替换，其中 x_2 是 9 年产量的均值.

1984	1985	1986	1987	1988	1989	1990	1991	1992
748.2	754.1	734.8	780.4	855.5	861.6	864.2	971.5	980.3

2. 一名学生对半径为 r 的圆 $x^2+y^2\leqslant r^2$ 内的整点[x，y]的个数 w 感兴趣(图 24.6). 他计算了一些 r 对应的 w：

r	1	2	3	4	5	6	7	8	9
w	5	13	29	45	81	113	149	197	253

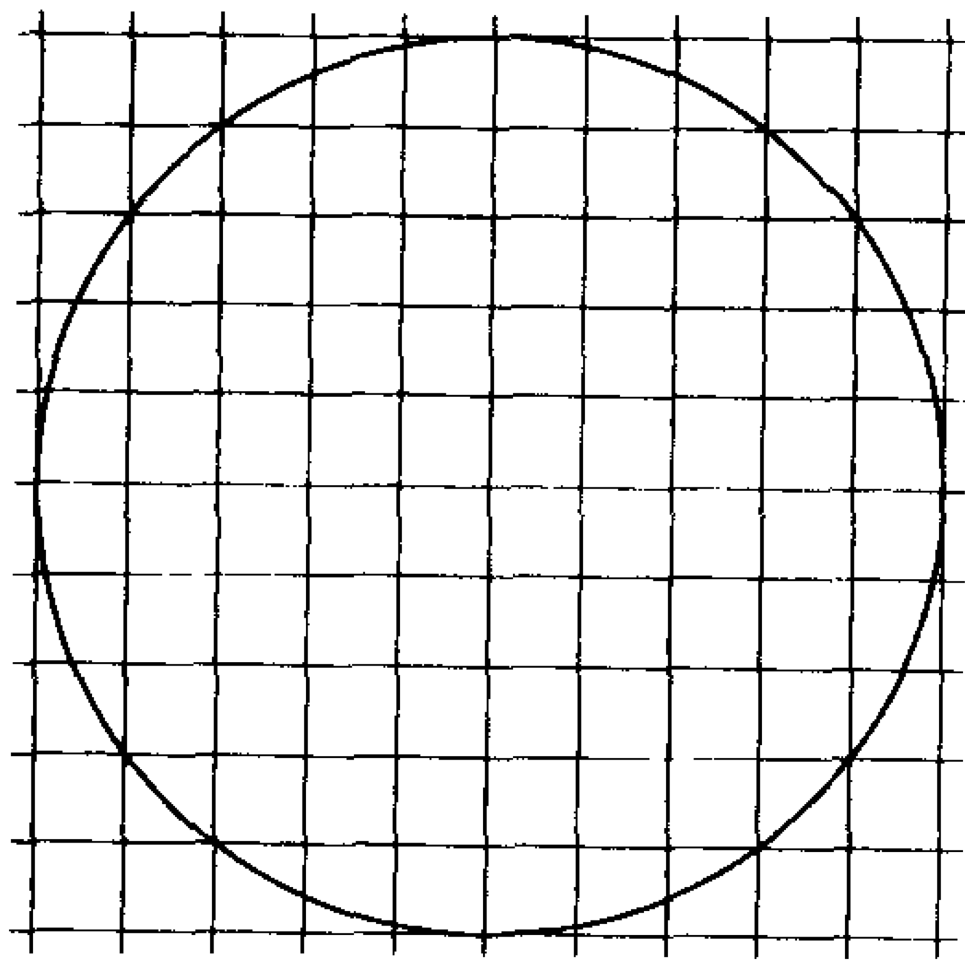


图 24.6 半径为 5 圆盘内的 81 个整点

这个学生希望用简单公式 $w=ar+b$(a，b 为常数)来近似 w. 但是，你觉得用圆盘的面积来近似更好，因此形如 $w=ar^2$ 的最佳 l^2 拟合对上面的数而言可能更好.

255

对两个模型 $w=ar+b$，$w=ar^2$ 分别计算最佳 l^2 拟合(最小二乘拟合)，说明哪个模型更好. 将两个最优值与下面的值比较：

$$\sum_{i=1}^{9}(w_i-\pi i^2)^2.$$

3. 求解例 24.4 中的优化问题. 提示：用图解法.

4. 由于航班超量售票，因此有些定了座的乘客不能登机(称为“堵塞”). 航空公司之所以超量售票，是因为他们不敢肯定有多少定了座的乘客将前来登机. 下面是美国最大的 10 个航空公司国内直航航班在第 t 年的登机人数 x(第 2 行，单位：百万人)和堵塞人数 y(第 3 行，单位：千人)，数据来自于 http://www.bts.gov/btsprod/nts/Ch1_web/1-55.htm：

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
421	429	445	449	457	460	481	503	514	523
628	646	764	683	824	843	957	1 072	1 126	1 070

分别对 $p=1, 2, \infty$，用 1990—1998 年的数据和模型 $y=at+bx+c$ 计算最佳 l^p 拟合，并用这些拟合预测 1999 年的人数 1 070. 256

附录　数学规划导引

A1. 数学规划

在第 1 节中我们提到，数学规划就是在一个集合 S 上对于 n 个变量 x_1，…，x_n 的一个函数 f 求极小值(或极大值)：

$$f(x) \to \min, x \in S. \tag{A1.1}$$

其中，决策变量 x_1，…，x_n 也称为控制变量、计划变量或策略变量，将它们记成列向量 x，则可行域 S 是所有实数项列向量集合 R^n 的一个子集. 集合 S 通常由下面三种类型的约束系统给出：$g(x)=0$，$g(x)\geqslant 0$，$g(x)\leqslant 0$，其中 $g(x)$ 是 R^n 上的一个函数. 注意同一个集合 S 可以由不同的约束集合给出. 从理论上看，S 可以只由一种约束类型 $g(x)=0$ 或者 $g(x)\leqslant 0$ 给出. 如果使用罚函数方法，也可以去掉所有约束条件.

实值函数 f 称为目标函数或极小函数(minimand)，它总可以认为是线性函数. 也就是说，给定规划问题(A1.1)，我们可以引入一个新的变量 z，并考虑等价问题

$$z \to \min, x \in S, f(x) - z \leqslant 0, \tag{A1.2}$$

这个问题有 $n+1$ 个变量，目标函数是线性函数. 规划问题(A1.1)和(A1.2)的最优值是一样的.

在这种非常一般的情形下，我们并不能在理论上有很大的发展，但可以给出两个定义. 在 R^n 中，我们赋予如下欧氏范数：

$$|x| = \|x\|_2 = (x^{\mathrm{T}}x)^{1/2}.$$

定义 A1.3　*集合 S 中的一个点 a 称为局部最优点(local optimum point 或 local optimizer)，或相对最优点(relative optimum point)或局部最优解(local optimal solution)，如果存在 $\varepsilon>0$，使得我们增加约束 $|x-a|\leqslant\varepsilon$ 之后，a 是最优解.*

注　*如果用 $\|\|_1$ 和 $\|\|_\infty$ 范数代替 $\|\|_2$ 范数，则约束 $|x-a|\leqslant\varepsilon$ 等价于一组线性约束. $\|\|_2$ 范数的优点之一在于它是光滑的(所有任意阶偏导数都存在). 在定义 A1.3 中使用哪种范数无关紧要，因为对于 R^n 中的所有向量 e 都有*

$$\|e\|_2/n^{1/2} \leqslant \|e\|_\infty \leqslant \|e\|_1 \leqslant \|e\|_2 n^{1/2}$$

例如，任何一个最优解(这时可以称为全局最优点)同时也是局部最优点. 在线性规划中，反过来也是对的(参见下面的附录 A3).

定义 A1.4 集合 S 中点 a 处的一个可行方向(feasible direction)是满足下列性质的非零向量 b：存在 $\varepsilon_0>0$，使得对于满足 $0\leqslant\varepsilon\leqslant\varepsilon_0$ 的任意 ε，都有 $a+b\varepsilon$ 属于集合 S.

注意这个定义与目标函数没有关系. 当 $S=R^n$ 时，每个点处的每个方向都是可行的.

数学规划的很多求解方法都需要在一个点处选择一个可行方向，使得从这个点沿这个方向移动一个小的步长时，目标函数的值可以得到改善. 然后采用线搜索方法确定沿这个方向移动多远，得到迭代过程的下一个点. 在点 x 处选择可行方向时，可能要用到目标函数和约束函数在点 x 处的值，以及这些函数的导数值. 有些迭代方法还要用到前面迭代点的值，这就需要从一个以上的初始点开始迭代.

在并行规划和遗传规划中，“时间”可能比序列“0，1，…”更复杂. 有时，可能完全任意或随机选择下一个点，否则我们得到的是确定型算法.

注意数学规划的迭代过程通常不能求出精确的最优解. 此外，得到的最后答案并不总是可行的. 原因之一在于，即使问题中所有的数据都是有理数，精确的答案仍可能会包括无理数.

从这个方面来说，线性规划是一个例外，因为单纯形法可以保证在有限步旋转后得到精确解. 但是，假设这个解是包括 10 000 000 000 位数字的一个整数，我们真的能算出这些数字吗？即使能够算出来，我们能使用这个答案吗？花大量时间算出这个精确答案值得吗？如果问题中所给的数据本身就是不精确的，将答案算到很多位有效数字是否有意义呢？ 258

带 10^{10} 位有效数字的答案看起来可能有些牵强，因为一个线性规划问题的最优解是一个线性方程组的唯一解(一旦知道基之后)，而这样的方程组在很多年以前就可以求解了. 然而在实际中，这样的方程组通常是求近似解.

有些迭代方法会产生 R^n 的一个无穷的序列

$$x^{(t)}=[x_1^{(t)},\cdots,x_n^{(t)}],t=0,1,\cdots$$

而有些方法会增加各种判停规则. 简单的判停规则是：经过一定次数的迭代或一定时间的计算以后，停止迭代. 如果对于所有的 $s>t$ 有 $x^{(t)}=x_s$，则迭代自然停止. 如果能够容易地判断一个解是否最优(或局部最优)，一个很好的判停规则是当 $x^{(t)}$ 是最优解(或相应地，是局部最优解)时，停止迭代. 有的判停规则是指定一个很小的正数 ε，称为*公差*(tolerance)，例如，当 $f(x^{(t)})$ 与最优值的偏差不超过 ε，或者 $x^{(t)}$ 与最优解的偏差不超过 ε 时，停止迭代(假设能够容易地判断与最优值或最优解的偏差). 也有些方法是当 $|x^{(t+1)}-x^{(t)}|\leqslant\varepsilon$ 或 $|f(x^{(t+1)})-f(x^{(t)})|\leqslant\varepsilon$ 时，停止迭代.

一种迭代方法通常作用于一类数学规划. 例如，使用导数的方法要假设导数存在. 如果我们对函数 $f(x)$ 一无所知，只知道序列 $x^{(t)}$ 不能穷举 S 以及

$f(x)$的相应值 $f(x^{(t)})$，那么我们就不知道离最优值有多远. 为保证一种方法能收敛或者能估计它离最优值有多远，我们需要为相应类的数学规划加上更多的限制. 对于 $f(x)$，一个典型的限制是带一个利普希茨常数 K 的利普希茨条件：

$$\text{对任意的 } x,y \text{ 有 } |f(x)-f(y)| \leqslant K|x-y|. \tag{A1.5}$$

这条件意味着 $f(x)$是连续的. 当梯度

$$f'(x) = \nabla f(x) = [\partial f(x)/\partial x_1, \cdots, \partial f(x)/\partial x_n] \tag{A1.6}$$

存在而且对任给 x 都连续时，(A1.5)等价于：对任给 x，$|f'(x)| \leqslant K$ 成立.

另一个有用的条件是 f 的凸性，它保证函数连续且局部最优解是(全局)最优解. 任何仿射函数 $f(x)=cx+d$ 都是凸的、光滑的，并且满足利普希茨条件，利普希茨常数 $K=|f'(x)|=|c|$.

259

求解(A1.1)迭代方法的一个理想性质是下降性质：

$$f(x^{(t+1)}) < f(x^{(t)}),\text{除非 } x^{(t+1)} = x^{(t)}$$

或者 $f(x)$的强下降性质(strong descent property)：当 $0<\alpha\leqslant 1$ 且 $x^{(t+1)} \neq x^{(t)}$ 时，$f((1-\alpha)x^{(t)}+\alpha x^{(t+1)})$是减函数.

单纯形方法就具有强下降性质.

我们再次探讨在目标函数和约束函数都可微的情况下一个点 x 成为最优解的条件. 首先考虑无约束最优化(即可行域 $S=R^n$)，那么一个点 $x=x^*=[x_1^*, \cdots, x_n^*]^T$ 是局部最优当且仅当它是临界点(critical)或驻点(stationary)(即梯度 $f'(x^*)=0$).

现在考虑 S 是可微函数 g_i 构成的一组约束条件 $g_i(x)=0$，$i=1$，…，k. 众所周知，此时如果 x^* 是局部最优解，那么梯度向量

$$\nabla f(x^*), \nabla g_1(x^*), \cdots, \nabla g_k(x^*) \tag{A1.7}$$

都线性相关.

在正则情况下，当$\nabla g_1(x^*)$，…，$\nabla g_k(x^*)$线性无关时，条件(A1.7)可以写成

$$\nabla f(x^*) = \sum_{i=1}^{k} \lambda_i \nabla g_i(x^*), \tag{A1.8}$$

其中系数 λ_i 称为拉格朗日乘子(Lagrange multiplier).

现在假设 S 是可微函数 g_i 构成的一组约束条件

$$g_i(x) = 0 \text{ 对 } i=1,\cdots,k \text{ 且 } g_i(x) \leqslant 0 \text{ 对 } i=k+1,\cdots,l \tag{A1.9}$$

我们称一个可行解 x 正则的(regular)，如果满足 $g_i(x)=0$ 的梯度$\nabla g_i(x)$(也就是有效梯度，active gradient)线性无关. 以下是一个正则可行解 x^* 成为局部极小的 Karush-Kuhn-Tucker(KKT)条件：

$$-\nabla f(x^*) = \sum_{i=1}^{l} \lambda_i \nabla g_i(x), \lambda_i g_i(z^*) = 0 \, \forall i, \lambda_i \geq 0 \text{ 对 } i > k.$$

条件 $\lambda_i g_i(z^*)=0 \, \forall i$ 意味着只涉及有效梯度. KKT 条件是关于 λ_i 的一组线性约束，且在$l=k$时就成了线性方程组(A1.8). 事实上由线性规划的对偶定理和隐函数存在定理可以知道 KKT 条件是必要的. 如果约束是线性的，就可以放弃正则性条件(condition of regularity). (当$l\leqslant 1$时也成立.)另外，如果 $f(x)$是凸的，KKT 条件就是 x^* 为最优解的充分条件(参看下面的附录 A3).

260

A2. 单变量规划

这是带有一个决策变量的数学规划(附录 A1 中的 $n=1$). 可行域经常是一段区间，一条射线或是整个实轴 R(无约束最优化). 我们有很多方法可以将带有多个变量的数学规划(多变量规划)归结为单变量规划或者直接利用单变量规划. 例如，我们可以尝试用“空间填充”曲线来近似 R^n 中的可行域 S，或者在一点选择一个可行方向，然后在这个方向上作直线搜索找到下一个点.

直接搜索法

当导数很难计算或者不存在时就用这种方法，然而这与使用导数的方法没有严格的区别，因为导数可以用 $f(x)$的值来近似.

一种在区间 $a\leqslant x\leqslant b$ 上最小化函数 $f(x)$的简单想法是：选择一个大的自然数 N，计算

$$f(a+i(b-a)/N) \text{ 对 } i=0,1,\cdots,N,$$

并找最小值. 当 $f(x)$满足利普希茨条件且利普希茨常数为 K 时，这个值与最优值的差不超过 $K/(2N)$. 更精细的方法是基于网格法(grid-based method)，需要将一些区间 $a+i(b-a)/N\leqslant x\leqslant a+(i+1)(b-a)/N$ 细化成更小的区间以搜索更好的解.

我们也可以在区间上任意选择 N 个点来计算 $f(x)$. 当 $f(x)$满足利普希茨条件时，若$N\to\infty$，$f(x)$在 N 个点的最小值就以概率 1 收敛于最优值.

我们称在一个在区间 $a\leqslant x\leqslant b$ 上的函数 $f(x)$是单峰的(unimodal)，如果它连续且只有一个局部最优解 x^*.

单峰函数最小值的斐波那契搜索法(Fibonacci search)要用到斐波那契序列 F_t(见 22 节练习第 13 题). 固定一个数 N，并且我们想通过求目标函数在 N 个合理选定点上的值，将未知的最优解 x^* 限制在最小可能区间上. 当 $N=0$ 或 1 时，我们不能将 x^* 限制在更小的区间上. 当 $N=2$ 时，通过比较 $f((a+b)/2+\varepsilon)$和 $f((a+b)/2-\varepsilon)$的值，其中 $0<\varepsilon<(a-b)/2$，将 x^* 限制在区间$a\leqslant x\leqslant (a+b)/2+\varepsilon$ 或$(a+b)/2-\varepsilon\leqslant x\leqslant b$ 上(当 $f'((a+b)/2)$存在时，可以用二分法省掉 ε).

261

对任意 $N>2$，比较 $f(a+(1-F_{N-1}/F_N)(b-a))$ 和 $f(a+(b-a)F_{N-1}/F_N)$ 的值，并将区间 $a\leqslant x\leqslant b$ 缩减为子区间 $a\leqslant x\leqslant a+(b-a)F_{N-1}/F_N$ 或子区间 $a+(1-F_{N-1}/F_N)(b-a)\leqslant x\leqslant b$. 这两种情况下子区间的长度均为 $(b-a)F_{N-1}/F_N$. 经过 $N-2$ 步，我们将 x^* 限制在长度为 $2/F_N$ 的子区间上. 最后两步估算使得对任意小的 $\varepsilon>0$，区间长度缩减为 $(1+\varepsilon)/F_n$.

当 $N\to\infty$ 时，斐波那契搜索法就变成**黄金分割搜索**(search by golden section)，此时我们比较 $f(\alpha a+(1-\alpha)b)$ 和 $f((1-\alpha)a+\alpha b)$ 的值，并将区间 $a\leqslant x\leqslant b$ 缩减为 $a\leqslant x\leqslant(1-\alpha)a+\alpha b$ 或 $\alpha a+(1-\alpha)b\leqslant x\leqslant b$，其中 $\alpha=2/(1+\sqrt{5})\approx 0.618$. $1/\alpha$ 是**黄金分割比**(golden section ratio). 经过 k 步估算，x^* 的不确定区间就被缩减为长度为 $(b-a)\alpha^{k-1}$ 的子区间.

对单峰函数，黄金分割搜索产生一个收敛于最优解 x^* 的序列 $x_t=x^{(t)}$. 如果应用于任意连续函数，这种方法产生一个收敛于点 x^* 的序列 x_t 使得

或者 $x^*=a$，或者有无穷多个 t 满足 $x_t<x^*$，且 $f(x_t)\geqslant f(x^*)$；

或者 $x^*=b$，或者有无穷多个 t 满足 $x_t<x^*$，且 $f(x_t)\geqslant f(x^*)$.

如果 $f'(x^*)$ 存在并且 $x^*\neq a, b$，则这样的点 x^* 就是临界点.

将可行区间分裂成更小的区间使我们能找到更多这样的点.

如果我们能够利用目标函数额外的信息，就可以设计出更好的方法. 例如，如果函数可微则会有帮助，此时如果 $f'(a)>0$，则 a 是局部最优的；如果 $f'(b)<0$，则 b 是局部最优的；而中间的点 $c(a<c<b)$ 是局部最优当且仅当它是临界点(即 $f'(c)=0$). 如果 $f(x)$ 是凸的或者如果 $F''(c)$ 存在且是正的，则临界点 c 是局部最小点.

二分法

这是用导数 $f'(x)$ 寻找局部极小点的方法. 假设 $f'(x)$ 在区间 $a\leqslant x\leqslant b$ 上存在而且 $f'(a)\leqslant 0\leqslant f'(b)$(否则我们就会有一个局部最优解). 我们令 $a_0=a$，$b_0=b$，$x_0=(a_0+b_0)/2$.

给定 a_t，b_t 和 $x_t=(a_t+b_t)/2$，我们定义 a_{t+1}，b_{t+1} 和 $x_{t+1}=(a_{t+1}+b_{t+1})/2$ 如下：如果 $f'(x_t)>0$，则 $a_{t+1}=a_t$，$b_{t+1}=x_t$，否则 $a_{t+1}=x_t$，$b_{t+1}=b_t$. 262

注意 $f'(a_t)\leqslant 0\leqslant f'(b_t)$，$a_{t+1}\leqslant a_t<b_{t+1}\leqslant b_t$，而且 $b_t-a_t=(b-a)/2^t$，对 $t=0, 1, \cdots$，$x_t=(a_t+b_t)/2$ 收敛，且 $x_t\to x^*$. 如果 $f'(x)$ 在 $x=x^*$ 连续则有 $f'(x^*)=0$，因此 x^* 是很好的局部最优解的候选点. 例如，如果对于接近 x^* 的 x 有 $f'(x)(x-x^*)\geqslant 0$，那么它就是局部最优解.

为了寻找一个以上的局部最优解，我们将起始区间 $a\leqslant x\leqslant b$ 分裂成许多小区间，然后对每一个小区间应用二分法.

如果导数不存在或者很难计算，就应用**二分搜索法**(dichotomous search)，其中通过估算 $f(c-\varepsilon)$ 和 $f(c+\varepsilon)$(ε 是小正数)来代替计算 $f'(c)$. 然而，黄金分

割法更加有效，因为它用两次估算将不确定区间缩小了$(1+\sqrt{5})^2/4\approx 2.61803$倍.

在无约束规划($S=R$)的情况下，我们可以不关心那些满足$|x|>M$的解x，其中M是一个很大的数. 这样我们得到一个S是有界区间的规划，因此我们可以应用二分法或其他设计出的方法做区间上的最优化. 我们也可以用一系列有限区间来覆盖R，但在无约束的情况下有更简单的方法.

梯度法

令$S=R$，$f'(x)$存在且对任意x是连续的. 我们从初始点x_0出发. 给定一个点x_t，梯度法中下一个点是$x_{t+1}=x_t-f'(x_t)$. 一种改进的梯度法如下给出：

$$x_{t+1} = x_t - \alpha f'(x_t), \tag{A2.1}$$

其中$\alpha>0$.

这种方法在驻点自然终止. 如果$f'(x)$满足利普希茨条件且利普希茨常数$K<2/\alpha$，则有

$$f(x_t) - f(x_{t+1}) \geqslant \alpha(1 - K\alpha/2)f'(x_t)^2 > 0.$$

因此我们就有$f(x)$的下降性质.

现在假设$f'(x)$满足利普希茨条件且利普希茨常数$K\leqslant 1/\alpha$，那么我们就有$f(x)$的强下降性质(见附录A1). 此外，我们还有$|f'(x)|$的强下降性质.

263

最后一条性质意味着序列x_t关于t在$f'(x_s)=0$的s之前是严格单调的. 当$f'(x_0)<0$ ($f'(x_0)>0$)时，或者序列在x_0的右侧(左侧)收敛于第一个驻点x^*，或者在右侧(左侧)没有驻点而且$x_t\to\infty$($x_t\to-\infty$).

如果附加以下条件

$$f'(x) - f'(y) \geqslant (x-y)K_2, \text{对所有 } x,y$$

对某个$K_2>0$成立(例如对任意x有$f''(x)\geqslant K_2$)，则函数$f(x)$是凸的(因此每一个驻点都是最优的)，而且我们可以估计收敛的速率：

$$f'(x_{t+1}) \leqslant f'(x_t)(K-K_2)/K \leqslant f'(x_0)(1-K_2/K)^{t+1}.$$

因此

$$|x_{t+1} - x_t| \leqslant \alpha |f'(x_0)| (1-K_2/K)^t$$

并且

$$|x_t - x^*| \leqslant \alpha |f'(x_0)| (1-K_2/K)^t K/K_2.$$

如果$f'(x)$的利普希茨常数不存在或者得不到，我们可以用*阻尼梯度法*(damped gradient method)

$$x_{t+1} = x_t - \alpha_t f'(x_t), \tag{A2.2}$$

其中*阻尼序列*(damping sequence)α_1，α_2，…，使得对任何t有$\alpha_t>0$，当$t\to\infty$时$\alpha_t\to 0$，而且$\sum\alpha_t=\infty$(见[V2]关于阻尼序列的讨论). 如果只有有限个临界点，

我们得到 x_t 收敛于一个临界点，∞或$-\infty$.

为找到多于一个的临界点，我们可以尝试不同的初始点. 例如，如果我们找到的临界点 x^* 不是局部最优解，我们可以用接近 x^* 的一点 y_0 再次开始，以使得 $f(y_0)<f(x^*)$. 人们已经找到更加复杂的以排除临界点的方式修正 $f(x)$的方法.

在 S 是有限区间 $a\leqslant x\leqslant b$ 的情况下，以 a 或 b 中更接近不可行的 x_{t+1}来替换 x_{t+1}，可以修正任何迭代方法.

为避免在梯度法中计算 $f'(x_t)$，我们可以用$(f(x_t)-f(x_{t-1}))/(x_t-x_{t-1})$替换它，并且从两个不同的初始点 x_0，x_1 开始. 264

牛顿法

比梯度法更复杂的方法是牛顿法，它作用于 $f(x)$为凸的区域里. 标准形式假设

$$f''(x) \text{ 存在，且对任意 } x \text{ 它都是正的.} \tag{A2.3}$$

然而，也有不使用一阶和/或二阶导数的形式. 我们从使用 $g(x)=f'(x)$但不用 $f''(x)$的形式开始.

假设导数 $g(x)=f'(x)$存在且连续，我们希望找到函数 $g(x)$的一个零点. 从两个初始点$x_0\neq x_1$开始，作一条经过点$(x_0, g(x_0))$，$(x_1, g(x_1))$的直线并截断这条直线，以此直线近似函数 $g(x)$且其相对于水平轴的斜率为

$$s=(g(x_1)-g(x_0))/(x_1-x_0),$$

因此我们取 $s=(0-g(x_1))/(x_2-x_1)$，从而得到下一个点 $x_2=x_1-g(x_1)/s$，其中 $s\neq 0$. 当$s=0$ 时，取 $x_2=(x_0+x_1)/2$. 然后用 x_1，x_2 而不是 x_0，x_1 来重复这一步骤，以此类推. 我们得到序列 x_t 希望它收敛于 $g(x)$的零点.

下面我们假设满足条件(A2.3)，因此目标函数 $f(x)$是凸的. 纯粹的找 $f(x)$极小点(即 $g(x)=f'(x)$的零点)的牛顿法以初始点 x_0 开始并且定义 x_t 如下：

$$x_{t+1}=x_t-g(x_t)/g'(x_t). \tag{A2.4}$$

为保证牛顿法(A2.4)对实解析函数收敛，我们还需要附加条件.

在 $g'(x)=f''(x)$满足利普希茨条件，利普希茨常数为 K_3，且对任意 x 满足 $|g(x)/g'(x)^2|\leqslant K_1$ 的情况下，我们有

$$|g(x_{t+1})|\leqslant (g(x)/g'(x))^2K_3/2\leqslant |g(x)|K_1K_3/2;$$

因此如果对任意 x 有 $K_1K_3/2=q<1$，则(A2.4)就收敛. 即在这一条件下，$g(x_{t+1})\leqslant g(x_0)q^{n+1}$. 如果对任意 x 有 $g'(x)\geqslant K_2>0$，而且从满足 $|g(x_0)|K_3/(2K_2^2)=q_0<1$ 的 x_0 开始，这种方法会收敛得很快(由于 $|g(x)|$ 的上升性质). 也就是我们有 $|g(x)|$ 的下降性质而且 $q_0|g(x_{t+1})|\leqslant(q_0|g(x_t)|)^2$，所以 265

$$q_0 \mid g(x_t) \mid \leqslant q_0^{2^t}$$

对任意 t 均成立.

一种修正的牛顿法是

$$x_{t+1} = x_t - \alpha g(x_t)/g'(x_t), \tag{A2.5}$$

其中 $\alpha>0$. 当对任意 x，y 有 $\alpha g'(y)/g'(x)\leqslant 1$ 时，这种方法收敛并且具有 $f(x)$ 的下降性质(见先前对梯度法的讨论). 有许多推广的拟牛顿法填补了梯度法和牛顿法之间的空白.

例如，阻尼牛顿法是

$$x_{t+1} = x_t - \alpha_t g(x_t)/g'(x_t),$$

其中 α_t 是 $\alpha_t\to 0$ 的序列(例如，$\alpha_t=1/t$).

牛顿法是基于二次型 $a_0+a_1x+a_2x^2$ 在单独一点或不同两点的信息来近似 $f(x)$的方法. 对于严格凸的二次型，它一步收敛. 更复杂的方法用到三次型的近似：

$$a_0+a_1x+a_2x^2+a_3x^3.$$

自协调函数

在[**NN**]中有关于修正牛顿法的重要分析. 它并不用(A2.5)中传统的在有界的 $f''(x)$上选择 α 并分析收敛性的方法，而是假设函数

$$f''(x)^{-1/2} \text{ 满足利普希茨条件且利普希茨常数为 } \kappa_2. \tag{A2.6}$$

在 $f'''(x)$存在的情况下，这个利普希茨条件等价于如下的在[NN]中引入并在其后的许多著作中使用的自协调(self-concordant)条件：

$$\mid f'''(x) \mid /f''(x)^{3/2} \leqslant 2\kappa_2 \text{ 对所有 } x. \tag{A2.6$'$}$$

266 在 $\kappa_2=0$ 的情况下，牛顿法一步终止得到最优解，所以我们假设 $\kappa_2>0$.

一方面，我们希望(A2.5)中的 α 较小使得我们有 $f(x)$的强下降性质以免陷入局部最小解. 既然假定对于任意 x 有 $f''(x)>0$，$f(x)$的强下降性质就等价于 $\mid g(x)\mid$ 的强下降性质. 另一方面，α 略小于必要值会减慢收敛速度. 那么，在条件(A2.6)下满足 $g(x_{t+1})=0$ 最小可能的 α 是多少呢？

利普希茨条件(A2.6)意味着

$$\mid g'(x)^{-1/2} - g'(y)^{-1/2} \mid \leqslant \kappa_2 \mid x-y \mid, \text{对所有 } x,y.$$

取 $x=x_t$ 且 $a= g'(x_t)^{-1/2}$，得

$$g'(y) \geqslant 1/(a+\kappa_2 \mid y-x_t \mid)^2, \text{对所有 } y \tag{A2.7}$$

以及

$$g'(y) \leqslant 1/(a-\kappa_2 \mid y-x_t \mid)^2, \text{当 } \mid y-x_t \mid < a/\kappa_2. \tag{A2.8}$$

令 $b=\mid g(x_t)\mid$，$g(x_t)<0$($g(x_t)>0$ 的情况类似). 在 $0\leqslant z<a/\kappa_2$ 条件下对

(A2.8)从 x_t 到 x_t+z 积分，得

$$g(x_t+z)\leqslant -b+1/(\kappa_2(a-\kappa_2 z))-1/(\kappa_2 a)\text{for}\,0\leqslant x< a/\kappa_2. \quad \text{(A2.9)}$$

在区间 $0\leqslant z<a/\kappa_2$ 上，(A2.8)式的右侧从 $-b$ 增加到 ∞ 并在 $z^*=a^2b/(1+ab\kappa_2)$ 取得零. 因此对于 $0\leqslant z\leqslant z^*$，有 $g(x_t+z)\leqslant 0$. 由等式

$$z^*=-\alpha g(x_t)/g'(x_t)=\alpha a^2 b$$

得

$$\alpha=\alpha_t=1/(1+ab\kappa_2)=1/(1+g'(x_t)^{-1/2}\,|\,g(x_t)\,|\,\kappa_2).$$

因此对这个 α 我们有(A2.5)的强下降性质(同一个 α 在 $g(x_t)>0$ 的情况下也成立).

现在估计 $|\,g(x_t)\,|$ 减小的速率. 对(A2.7)积分，得

$$g(x_t+x)\geqslant -b-1/(\kappa_2(a+\kappa_2 x)+1/(\kappa_2 a),\text{对任意的 } x\geqslant 0. \quad \text{(A2.10)}$$

267

取

$$x=x^*=a^2b/(1+ab\kappa_2),$$

我们得到

$$|\,g(x_{t+1})\,|\leqslant 2ab^2\kappa_2/(1+2ab\kappa_2)=|\,g(x_t)\,|\,(1-\alpha)/(1-\alpha/2). \quad \text{(A2.11)}$$

(在 $g(x_t)\geqslant 0$ 情况下也同样成立.)

现在假定

$$ab=|\,g(x_t)\,|\,/g'(x_t)^{1/2}$$

是有界的，即

$$|\,f'(x)\,|\leqslant \kappa_1 f''(x)^{1/2}\ \text{对所有 } x. \quad \text{(A2.12)}$$

令

$$\kappa=\kappa_1\kappa_2,\quad \alpha^*=1/(1+\kappa),\quad \beta^*=(1-\alpha^*)(1-\alpha^*/2)=2\kappa/(1+2\kappa). \quad \text{(A2.13)}$$

那么对任意 t 有 $\alpha^*\leqslant\alpha_t\leqslant 1$，且

$$|\,g(x_{t+1})\,|\leqslant|\,g(x_t)\,|\,\beta\leqslant|\,g(x_t)\,|\,\beta^*$$

对所有 t 都成立，因此当 $t\to\infty$ 时

$$|\,g(x_t)\,|\leqslant|\,g(x_0)\,|\,\beta^{*t}\to 0.$$

我们有 $|\,g(x)\,|$ 的强下降性质，意味着 $t\to\infty$ 时 $x_t\to x^*$ 单调收敛，其中 x^* 是一个数或者 $\pm\infty$. 当 $|\,x^*\,|<\infty$ 时，对于介于 x_0 和最小值 x^* 之间的所有 x，我们有 $g'(x)\geqslant K_2>0$. $g'(x)$ 的这个下界意味着

$$|\,x_t-x^*\,|\leqslant|\,g(x_t)\,|\,/K_2\leqslant|\,g(x_0)\,|\,\beta^t/K_2.$$

而且

$$|g(x_{t+1})| \leqslant |g(x_t)ab| \leqslant cg(x_t)^2,$$

其中对所有的 t 有 $c=\kappa_2/K_2^{1/2}$. 一旦有某个 s 使得 $|cg(x_s)|=q<1$，我们就有

$$|cg(x_{s+t})| \leqslant q^{2^t}$$

[268] 对所有的 t 都成立.

例 A2.14 令 $f(x)=-c_1\log(x)-c_2\log(1-x)$，其中 c_1，$c_2>0$. 在区间 $0<x<1$ 上我们有

$$f''(x) \geqslant K_2 = (c_1^{1/3}+c_2^{1/3})^3$$

以及(A2.6)(其中 $\kappa_2=(\min[c_1, c_2])^{-1/2}$)和(A2.12)(其中 $\kappa_1=(\max[c_1, c_2])^{1/2}$)，从而

$$\kappa = \kappa_1\kappa_2 = (\max[c_1,c_2]/\min[c_1,c_2])^{1/2}.$$

如果我们以区间中的 x_0 开始，并且用 $\alpha=\alpha^*=1/(1+\kappa)$的修正牛顿法，那么根据强下降性质，我们得到的解会保持在区间中，并且有 $|f'(x_t)| \leqslant |f'(x_t)|\beta^{*t}$和 $|x_t-x^*| \leqslant |f'(x_t)|\beta^t/K_2$，其中

$$\beta^* = 2\kappa/(1+2\kappa) = (1-\alpha)/(1-\alpha/2) < 1.$$

应用可变步长

$$\alpha = \alpha_t = 1/(1+g'(x_t)^{-1/2}|g(x_t)|\kappa_2) \geqslant \alpha^*. \tag{A2.15}$$

在保持 $f(x)$的强下降性质的同时我们得到更快的收敛性(形式为 $|x_t-x^*| \leqslant cq^{2^t}$，其中 $0\leqslant q<1$).

■

例 A2.16 这是上例的推广.

令

$$f(x)=-\sum_{i=1}^{m} c_i\log(b_i-a_ix),$$

其中所有 $c_i>0$. 这个函数定义在由约束 $a_ix<b_i$ 给出的集合 S 上. 集合 S 可以如下给出：$a<x<b$，其中 a 是一个数或 $-\infty$，b 是一个数或∞. 现在假设 $a<b$(即 S 非空).

对 S 中所有 x，我们有(A2.12)(其中 $\kappa_1=(\sum c_i)^{1/2}$)以及(A2.6)(其中 $\kappa_2=\min(c_i)^{-1/2}$).

在所有 $c_i=1$ 的特殊情况下，$\kappa_1=m^{1/2}$，$\kappa_2=1$ 和 $\kappa=m^{1/2}$，所以我们可以取 $\alpha^*=1/(1+m^{1/2})$.

当 S 有界时，对 S 中所有 x 有 $f''(x) \geqslant K_2>0$. 如果 S 无界，我们就得不到这样的结果. 当 $a=-\infty$且 $f'(x_0)>0$ 时，我们有 $\min=-\infty$且 $x_t\to-\infty$. 当 $b=\infty$且 $f'(x_0)<0$ 时，我们有 $\min=-\infty$且 $x_t\to\infty$. 所有其他情况下，我们有

$$|x_t-x^*| \leqslant c_1(2m^{1/2}/(1+2m^{1/2}))^t.$$

如果对固定的

$$\alpha = \alpha^* = 1/(1+m^{1/2})$$

应用修正牛顿法及可变步长(A2.15)，我们会有更快的收敛性. ■ 269

A3. 凸规划与二次规划

这部分是在凸集 S 上极小化凸函数 $f(x)$，凸规划推广了线性规划.

定理 A3.1 对任何凸规划，任何局部最优解都是(全局)最优的.

证明 (参见[B 1]，[FV]). 令 x^* 是局部最优解且 y 是可行解，我们要证明 $f(x^*)\leqslant f(y)$. 假设 $f(x^*)>f(y)$，既然 S 是凸的，$z=(1-a)x^*+ay$ 就是满足 $0\leqslant a\leqslant 1$ 的可行解，且对$0<a\leqslant 1$，有 $f(z)\leqslant(1-a)f(x^*)+af(y)<f(x^*)$. 对于接近 0 的 a，z 接近 x^*，这与 x^* 是局部最优矛盾. ■

如有必要，可以用(A1.2)的技巧将 $f(x)$化为线性型同时保持可行集是凸的. 集合 S 通常由一个或若干个形如 $g(x)\leqslant 0$ 的约束给定，其中 g 是凸函数. 除了凸外，这样的集合 S 还是闭的(即包含它的极限点). 相反地，任何闭集都可由一个约束 $g(x)\leqslant 0$ 给定，其中 $g(x)$是凸函数. 在 S 由凸集系统 $g_i(x)\leqslant 0$，$i=1,\cdots,m$ 给定的情况下，这样的凸函数 $g(x)$可以写为 $g(x)=\max[g_1(x),\cdots,g_m(x)]$.

一些凸集 S 由涉及拟凸函数(quasiconvex functions)的约束给定.

定义 A3.2 一个函数 $g(x)$称为是拟凸的(quasiconvex)，如果集合 $g(x)\leqslant c$ 对任何 c 都是凸的.

等价地，对任何 x，y 以及区间 $0\leqslant a\leqslant 1$ 中的 a，有 $g(ax+(1-a)y)\leqslant\max(f(x),f(y))$. 每一个凸函数都是拟凸的. 单变量 x 的函数 $g(x)$是拟凸的，如果它是单峰的.

注意一些拟凸约束 $g_i(x)\leqslant 0$，$i=1,\cdots,m$ 可以由单一拟凸约束 $g(x)=\max[g_1(x),\cdots,g_m(x)]$替代.

在 $S=R^n$ 且函数 f 可微的情况下，x 是最优的当且仅当$\nabla f(x)=0$. 一般地，我们有如下结果.

定理 A3.3 令 x^* 是如下凸规划的一个可行解：

$$f(x)\to\min, g_i(x)\leqslant 0 \text{ 对 } i=1,\cdots,m, \qquad (A3.4)$$

其中 $f(x)$是凸的且在 x^* 点可微，并且所有的 $g_i(x)$是拟凸的且在 x^* 点可微. 如果 x^* 满足 KKT 条件(参见前面的附录 A1)，那么 x^* 是最优的. 270

证明 令

$$-\nabla f(x^*)=\sum_{i=1}^{m}\lambda_i\nabla g_i(x),$$

其中对任意 i 有 $\lambda_i g_i(z^*)\geqslant 0$ 且当 $g_i(x^*)=0$ 时 $\lambda_i=0$.

我们要证明对任意可行解 x' 有 $f(x^*)\leqslant f(x')$. 既然对任意有效约束有 $g_i(x)\leqslant 0=g_i(x^*)$且 $g_i(x)$限制在连结 x^* 和 x' 的线段上是拟凸的，因此我们有 $\nabla g_i(x^*)(x'-x^*)\leqslant 0$. 于是

$$\nabla f(x^*)(x'-x^*)=-\sum_{i=1}^{m}\lambda_i \nabla g_i(x)(x'-x^*)\geqslant 0.$$

既然 $f(x)$在线段上是凸的，我们有 $f(x^*)\leqslant f(x')$. ■

推论 A3.5 *设 x^* 是如下凸规划的正则可行解 $f(x)\to\min$, $g_i(x)\leqslant 0$, $i=1,\cdots,k$, 其中 $f(x)$是凸的且在 x^* 点可微，且所有 $g_i(x)$是拟凸的且在 x^* 点可微，那么 x^* 是最优的当且仅当它满足 KKT 条件.*

在 $f(x)$二阶连续可微的情况下，$f(x)$是凸的当且仅当它的黑塞矩阵(Hessian)，即

$$\nabla^2 f(x)=[\nabla^2_{i,j}f(x)]=f''(x)=\left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right] \tag{A3.6}$$

$=Q$ 是半正定的(positive semidefinite)(即对 R^n 中所有的 y, $y^{\mathrm{T}}Qy\geqslant 0$). 等价地，二次型 $y^{\mathrm{T}}Qy\geqslant 0$ 是线性型的平方和.

传统上，二次规划(quadratic programming)是凸规划的一部分. 它是关于极小化线性型平方和与一个仿射函数之和，并带有线性约束的一个有限集合. 目标函数可以写成的 $f(x)=d+cx+x^{\mathrm{T}}ax$ 形式，其中是 x 是 n 个决策变量的列向量，d 是一给定的数，c 是给定 n 个数值的行向量，a 是 $n\times n$ 的对称半正定矩阵. 既然是凸函数的和，$f(x)$是凸的，因此二次规划是凸规划的特殊情况. 当二次规划有界时，它有最优解. 这对凸规划也成立，正如单变量无约束的例子 $e^x\to\min$ 所展示的那样. 当 $a=0$ 时，二次规划就成为线性规划.

271

当所有的约束都是等式时，可以通过求解线性方程组找到一个最优解. 也就是，我们可以求解约束系统，并除去变量将问题简化为无约束的二次规划问题. 然后可以令所有的偏导数为零找到所有的最优解. 无约束凸规划常使用的方法是牛顿法和拟牛顿法，参见下一节. 在二次型的情况下，牛顿法一步给出精确最优解.

在有约束的情况下，经常使用内点法，参见附录 A5. 许多约束优化方法在凸规划的情况下非常有效，参见附录 A4. 但是对于非凸规划，这些方法常常是启发式的(heuristic)，即用它们可以寻找解，但不能保证一定找得到. 与此形成对照的是，对于凸规划，这些方法有很好的收敛性质. 此外，如果对黑塞矩阵附加额外的限制，我们就能估计收敛速率，参见附录 A2 和附录 A4.

求解凸规划的一个方法是用分段线性函数近似凸目标函数(即用有限个仿射函数中最大者). 我们也可以用线性约束的有限系统近似可行域 S. 当 S 是由约束 $g_i(x)\leqslant 0$(其中 $g_i(x)$是凸的)给定时，它可以用分段线性函数近似每一个 $g_i(x)$来取得，关键是包含分段线性函数$f(x)$ 和$g_i(x)$的凸规划(A3.4)等价于线

性规划．单纯形法有一些特殊的修正可以处理这样的线性规划，其中一些用于可变大小的情况，必要时可以添加行和列，同时去掉多余的行和列．

线性规划中的对偶性已经扩展到凸规划中．例如，对于二次规划

$$x^{\mathrm{T}}Qx + cx \to \min, Ax \leqslant b, x \leqslant 0,$$

其中 $Q=Q^{\mathrm{T}}$，根据[D]，它的对偶是

$$-y^{t}Qy + bu \to \max, u \geqslant 0, A^{\mathrm{T}}u - 2Qy \leqslant c.$$

最后，有一些技巧可将非凸规划转化为凸规划．例如，考虑规划 $f(x)\to\min$，其中 $f(x)$ 二次可微．如果对某个 $K\geqslant 0$ 和 R^n 中所有的 x，u 有 $K(f'(x)u)^2 \geqslant -u^{\mathrm{T}} f''(x)u$（或 $K(f'(x)u)^2 \geqslant -u^{\mathrm{T}} f''(x)uf(x)$），那么我们的规划就和凸规划有相同的最优解 $\mathrm{sign}(f(x)) \mid f(x) \mid^{K+1} \to \min$（或 $e^{Kf(x)} \to \min$）．

272

A4．多变量规划

单变量最优化方法已经以多种方式推广或应用在多变量最优化中．当可行域是整个空间 R^n 时一些方法会简单得多，因此我们先以无约束最优化开始，然后讨论有约束的情形．

坐标下降法

从初始点 $x^{(0)}=[x_1^{(0)}, \cdots, x_n^{(0)}]$ 开始．为找到下一个点，我们选择一个下降坐标(descent coordinate)并且保持其他变量不变而最优化 $f(x)$ 在这个变量下的值．然后我们用其他坐标继续这样做，得到序列 $x^{(0)}$，$x^{(1)}$，…，满足 $f(x^{(t+1)}) \leqslant f(x^{(t)})$．一种下降坐标的选择可能是循环的：

$$x_1, x_2, \cdots, x_n;\ x_1, x_2, \cdots.$$

如果对某个 t 有 $f(x^{(t+n)})=f(x^{(t)})$，且 $f'(x)$ 存在并在点 $x=x^{(t)}$ 连续，那么 $f'(x^{(t)})=0$.

当 $f(x)$ 的梯度 $f'(x)$ 可以得到时，坐标下降法是一个更好的选择．这种方法比梯度法简单（见下面），但是收敛性通常比较糟糕．

基于网格法

基于网格法是在经过细分的网格上，采用不需要导数的搜索方法搜索局部最优解．见[CP1]．

基于单纯形法

基于单纯形法构造一个 R^n 中 $n+1$ 个不同点的进化模式，这些点可被看作是一个单纯形的顶点（在 $n=1$ 的情况下我们有一对进化点并且单纯形是一个区间，参见附录 A2）．在某些方法中，通过应用简单法则得到下一个单纯形，诸如从 $f(x)$ 最大值的顶点反射出来并压缩到 $f(x)$ 最小值的顶点．

在其他方法中，进化更加复杂，主要的思想是在每一步迭代中目标函数值最

好有严格的改进(参见[K2],[SMY]).

梯度法(也称为最速下降法)

我们希望从初始点 x_0 开始找到连续可微函数 $f(x)$的临界点(使之最小). 下一个点 $x^{(t+1)}=F(y, t)$取决于前一个点 $x^{(t)}=y$:

273

$$F(y,t) = y - \alpha_t f'(y)^{\mathrm{T}}, \tag{A4.1}$$

这里 α_t 称为步长(step size),而 $f'(x)=\nabla f(x)$是梯度(见(A1.6)).

这种方法当遇到临界点 y 时自然终止此时 $F(y, t)=y$. 在此之前,对足够小的 α_t,每一步都改进我们的目标函数(即我们有强下降性质).

步长是 $b_t=|x^{(t+1)}-x^{(t)}|=|\alpha_t f'(y)|$. 因为要收敛于一点我们希望 $b_t\to 0$(这对应于模拟退火中的冷却). 为了找到一个可能远离初始点的最优解,我们希望$\sum 1/b_t=\infty$,或者尝试多个不同的初始点. 如果 $x^{(t)}$ 收敛到点 z,那么在 $f'(x)$连续的条件下,$f'(z)=0$.

α_t 的理想选择是使单变量极小化问题 $f(y-\alpha f'(y)^{\mathrm{T}})\to\min$(其中 $y=x^{(t)}$)最优的 α 的值. 理想线搜索可以在某些情况下精确地完成(例如,$f(x)$是 2 次多项式). 在一般情况下,线搜索是近似完成的(见附录 A2). 例如,我们可以令 $\alpha_t=1/t$ 或 $b_t=1/t$(参见 Brown 方法).

当 $f(x)$不可微或者它的导数难于计算时,可以用任意一个使 $f(x)$下降的方向替代梯度法中的$-f'(y)$(可行方向法). 另一方面,当 $f(x)$的二阶导数可得到时,更倾向于使用牛顿法(下面讨论).

例如,令 $f(x)=x^{\mathrm{T}}Qx+cx$,其中 Q 是对称正定矩阵(即对 R^n 中任何 $z\neq 0$,有 $z^{\mathrm{T}}Qz>0$),那么就有唯一的最优解 $x^*=-Q^{-1}c^{\mathrm{T}}/2$,而且对某个 α 和 $0<\beta<1$ 带有理想线搜索的梯度法给出了全局收敛的 $|x_t-x^*|\leqslant\alpha\beta^t$. (类似的结论在更一般的情况下也成立,如果 $f(x)$是凸函数,二阶导数连续,且当 R^n 中 x, y, z 满足 $|y|=|z|=1$ 时 $y^{\mathrm{T}}f''(x)y/z^{\mathrm{T}}f''(x)z$ 有界,见[L])此时牛顿法一步给出全局收敛解:$x_1=x^*$.

信赖域法

给定一点 w,为找到下一个点 $F(w)$,我们极小化价值函数(merit function) $\tilde{f}(x)$,它在信赖域(trust region)S'上近似 $f(x)$. 我们在 S'上信赖我们的近似函数. 见[CGT].

例如,可微函数 $f(x)$在点 w 的线性近似是关于 x 的仿射函数

274

$$f(w) + f'(w)(x - w), \tag{A4.2}$$

令 S'为球 $|x-w|\leqslant\alpha_t|f'(w)|$,可得到上述(A4.1)的梯度法.

应用线性近似但 $S'=S$ 由线性约束的一个有限集合给定,可得到凹规划中找局部最优解的方法,见下.

在无约束的情况下,取线性近似为 $\tilde{f}(x)$,且 S'由$(x-w)^{\mathrm{T}}Q(x-w)\leqslant 1$ 给

定(Q是对称正定矩阵)，我们得到变尺度法(variable metric method).

再次应用线性近似但根据约束的不同换成不同的信赖域，我们得到仿射尺度法(见附录A5).

现在假设黑塞矩阵$f''(x)$(见(A3.6))存在. $f(x)$的二次近似(quadratic approximation)

$$\tilde{f}(x) = f(w) + f'(w)(x-w) + (x-w)^{\mathrm{T}} f''(w)(x-w)/2 \quad \text{(A4.3)}$$

在w点是关于x的次数不大于2的多项式. 假设矩阵$Q=f''(w)$正定，应用二次近似和$S'=R^n$，我们得到下面的牛顿法.

牛顿法

假设矩阵$f''(x)$存在且在所有的点x连续可逆. 牛顿法中的下一个点$F(w)$是二次近似$\tilde{f}(x)$的临界点(见(A4.3))，可以明确地计算为

$$F(w) = w - f''(w)^{-1} f'(w)^{\mathrm{T}}. \quad \text{(A4.4)}$$

注意到$F(w)$是$\tilde{f}(x)$的最小解当且仅当矩阵$f''(w)$正定，即$\tilde{f}(x)$是凸的.

为了能够收敛，通常对$f(x)$的二阶导数应用利普希茨条件，或者对$f(x)$的三阶导数加以限制. 例如，假设对R^n中任意满足$|d|=1$的x，d有$f''_d(x) \geqslant K_0 > 0$，且对$R^n$中任意满足$|d|=1$的$d$，$f''_d(x)$满足利普希茨常数$K_3$. 这里$f''_d$是$f$的方向导数，即$f$限制在直线$x=w+dz$上的二阶导数$\mathrm{d}^2/\mathrm{d}z^2$. 那么

$$|f'(F(w))| \leqslant |f'(F(w))|^2 K_3/(2K_0^2). \quad \text{(A4.5)}$$

这样，如果$f''(x^*)$正定并且我们的初始点离x^*足够近，牛顿法就很好地收敛于局部最小解x^*. 我们无需假设$f(x)$在R^n中的每一点都是凸的. 阻尼形式 [275]

$$F(w) = w - \alpha f''(w)^{-1} f'(w)^{\mathrm{T}} \quad \text{(A4.6)}$$

(其中$0<\alpha\leqslant 1$)可用于增加区域的收敛性.

对于比较大的n，矩阵$f''(w)^{-1}$通常是近似值，用它的前一个值作为起始点. 这引出了共轭方向法(conjugate direction method)和拟牛顿法(quasi-Newton Method)，它们填补了梯度法和牛顿法之间的空白(见[L, Ch. 8和9]).

自协调函数

假设$f''(x)$存在连续正定，且$f(x)$在任何直线上的限制满足条件(A2.6)，即

$$(e^{\mathrm{T}} f''(x)e)^{-1/2} u - (e^{\mathrm{T}} f''(x+e)e)^{-1/2} \leqslant \kappa_2, \text{对所有 } x, e \in R^n, e \neq 0. \quad \text{(A4.7)}$$

正如我们在附录A2中看到的，修正牛顿法(A4.6)对$f(x)$有强下降性质，只要

$$\alpha \leqslant \frac{1}{1+\kappa_2 |f'(w)u| |u| / (u^{\mathrm{T}} f''(w)u)^{1/2}},$$

其中 $u=-f''(x)^{-1}f'(w)^{\mathrm{T}}$ 是 w 点的牛顿方向，从而 $f'(w)u=-f'(w)f''(x)^{-1}f'(w)^{\mathrm{T}}<0$.

但是为使 $|f'(x)|$（而不是 $|f'(x)u/|u||$）减少得更多，我们希望一个更小的

$$\alpha \leqslant \alpha_t = \frac{1}{1+\kappa_2 |f'(w)|^{1/2} |u|^{-1/2}(-f'(w)^{\mathrm{T}} f''(w) u)^{-1/2}}, \tag{A4.8}$$

其中

$$\begin{aligned} 0 < -f'(w)^{\mathrm{T}} f''(w) u &= f'(w)^{\mathrm{T}} f''(w) f'(w) \\ &\leqslant (u^{\mathrm{T}} f''(w) u)^{1/2} (f'(w)^{\mathrm{T}} f''(w) f'(w))^{1/2}. \end{aligned}$$

对于这样的 α 我们有（假设像在[NN]中一样三阶导数存在且连续）

276

$$|f'(F(w))| \leqslant |f'(w)|(1-\alpha)/(1-\alpha/2).$$

现在引入(A2.12)的 n 维形式

$$f'(x) f''(x)^{-1} f'(x)^{\mathrm{T}} \leqslant \kappa_1 |f'(x)|^{1/2} f''(x)^{-1} f'(x)^{\mathrm{T}} |^{1/2} | \ \forall x \in R^n. \tag{A4.9}$$

在此条件下，令

$$\kappa = \kappa_1 \kappa_2, \alpha = 1/(1+\kappa), \beta = 2\kappa/(1+2\kappa) = (1-\alpha)/(1-\alpha/2).$$

那么对任意 t 有 $\alpha^* \leqslant \alpha_t$，而且方法(A4.6)有强下降性质，以及对于任意满足 $x^{(t)}=F^t x^{(0)}$ 的 t 有 $f'(x^{(t)}) \leqslant f'(x^{(0)})\beta^t$.

于是，$|x^{(t)}| \to \infty$，或者 $x^{(t)} \to x^*$（x^* 属于 R^n）. 如果增加条件

$$u^{\mathrm{T}} f''(x) u \geqslant K_2 |u|^2 \ \forall u \in R^n, \tag{A4.10}$$

其中 $K_2>0$，我们得到 $|x^{(t)}-x^*| \leqslant \beta^t/K_2$. 如同(A4.8)中一样取 $\alpha=\alpha_t \geqslant \alpha^*$，我们得到在 x^* 附近更快的收敛.

现在我们讨论有约束最优化问题，此时要想找到一个可行解都是困难的.

在某些情况下，对于可以让我们去掉那些约束和变量的一些变量，定义 S 的约束可被明确地解出. 例如，如果我们所有的约束都是线性方程，我们就可以除去所有的约束而得到无约束规划.

可以对一些迭代方法加以限制而用于有约束的情况，如有必要，步长保持为 F. 如果我们碰到边界，需要更加复杂的方法，比如梯度投影法（找最接近 $-f'(x)$ 的可行方向），以保持下一个点在 S 中（见下面的讨论和[L]）. 因此用障碍法避开边界是个好主意（见下一节）.

罚函数法

去除(A1.1)中的约束，广泛使用的方法涉及**罚函数**(penalty function)法. 这个函数 $p(x)$ 应该在 S 上为零而在其他地方为正. 因此(A1.1)的最优解是无约束规划 $p(x) \to \min$ 的最优解.

规划(A1.1)由无约束规划 P_k 的序列 $f(x)+c_kp(x)\to\min$ 替代，其中序列 $0<c_k\to\infty$ 是对违反约束行为的大惩罚.

假设 $f(x)$ 和 $p(x)$ 都连续且规划 P_k 有最优解 $x^{(k)}$(这是必然的，如果 $|x|\to\infty$ 时，$f(x)/p(x)\to 0$)，序列 $x^{(k)}$ 的每一个极限点都是规划 $f(x)\to\min$ 的最优解，其中 x 属于 S.

277

当 S 由约束(A1.8)给定时，罚函数可写为

$$p(x)=\sum_{i=1}^{k}g_i(x)^2+\sum_{i=k+1}^{l}(\max[0,g_i(x)])^2. \qquad \text{(A4.11)}$$

当所有的 $g_i(x)$ 都可微时，$p(x)$ 也可微.

注意同一个 S 可由不同的约束系统给定，这样产生了不同的罚函数. 为得到 $p(x)$ 使得 $f(x)+c_kp(x)$ 的最小化能由所选方法(比如牛顿法)有效地完成，就需要好的判断. 另一个需要判断的是如何调整求解 P_k 过程中的精确度，这涉及选择序列 c_k.

当规划(A3.4)中的 $g_i(x)$ 是凸函数时，函数 $p(x)$ 就成为凸函数 $p(x)=\sum_{i=1}^{m}(\max[0,\ g_i(x)])^2$，因此每个 P_k 都是凸规划. 此时，另一凸罚函数是 $p(x)=\max[0,\ g_1(x),\ \cdots,\ g_m(x)]$.

罚函数称为**精确的**(exact)，如果对某个数 c，无约束问题 $f(x)+cp(x)\to\min$ 的局部最小解也是原始规划(A1.1)的局部最小解. 搜索这样的函数会引出对偶性和 KKT 条件. 见[L]. 一些迭代方法将对偶性和罚函数的思想联接或合并在一起.

线性约束

假设 S 由线性约束的有限系统给定，如同线性规划中一样. 如果所有的约束都是线性方程，那么我们可以求解这个线性方程组. 如果没有可行解或只有一个可行解，那么问题结束. 否则，我们可以除去一些变量而得到等价的无约束最优化问题(带有更少数目的变量).

在线性规划情形中，一般地找到一个可行解与找到一个最优解一样困难. 一种方法是单纯形法. 椭球法可用于线性或凸规划，见[H1]. 当然，还有**非可行内点法**(infeasible interior method)，它与内点法很相似但却是外点法，同时为了产生一个可行解用到非可行解.

当 S 由线性约束给定而且 $f'(x)=\nabla f(x)$ 存在时，建议使用下面的线性规划方法.

278

给定可行解 w，极小化 $f(x)$ 在点 w 的线性近似(A4.1)，它是关于 x 的仿射函数，以便找到下一个点 $F(w)$. 注意当 $f(x)$ 凸时我们有强下降性质. 然而，如果最优解不在顶点，就需要在连结点 w 和 $F(w)$ 的线段上进行线搜索以便得到收敛. 下面是一个阻尼的例子：

$$x^{(t+1)} = x^{(t)} + (F(x^{(t)}) - x^{(t)})/t.$$

为找到初始点 $x^{(0)}$（或者发现规划不可行）我们可以在 S 上最小化任意线性型．当 $f'(x^{(t)})=0$ 时终止这一程序．在凸的情况下，显然，或者 $x^{(t)}$ 收敛于最优解或者 $|x^{(t)}| \to \infty$．一般情形下，序列的每一个极限点都是临界点．

在[FW]中带有理想线搜索的方法被建议用于二次规划．此时线搜索由牛顿法一步迭代完成，而且如果 S 非空 $f(x)$ 有下界，最优解存在．

如果可行解 y 的方向 d 可以改进 $f(x)$ 但不可行，我们怎么办？考虑对应于有效约束的线性系统 $Ax=b$，将 d 投射到子空间上，即用子空间中最接近 d 的向量 pd 替换 d，见 23 节．当 y 正则时（即 AA^{T} 可逆）矩阵 p 的形式是

$$p = I_n - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}A.$$

方向 pd 是可行的，而且如果 $pd \neq 0$ 它还可以改进 $f(x)$．在 $d=-f'(x)$ 情形下我们得到*梯度投影法*(gradient projection method)．

在目标函数是凹的（即 $-f(x)$ 是凸的）情形下，我们有角点原理，因此不需要阻尼．凹规划在一些应用[VCS]中出现．一旦找到了一个局部最优解，就可以应用附加的线性约束除去它（*割平面法*(cutting plane methods)，见[L，13 章]）．重复这一过程可以得到一个新的局部最优解．在有限步内得到一个最优解（如果存在的话）是可能的．见[HT]，[P]．

注意更加一般的可行域，尤其是凸的区域，可以由线性约束给定的区域来近似．这扩展了本小节所提方法可能应用的范围．

279

A5．内点法

给定 R^n 中的集合 S，S 中的一点 a 叫做*内点*，如果存在 $\varepsilon>0$ 使得球 $|x-a| \leqslant \varepsilon$ 包含在 S 中．

解数学规划 $f(x)\to\min$，$x\in S$ 的一种内点法从内点 $x^{(0)}$ 开始并生成收敛于最优解的内点序列 $x^{(0)}$，$x^{(1)}$，…（或者发现没有最优解）．在某些情况下这样的全局收敛可能只是奢望，若能得到如下的结果就可以：$f(x^{(t)})\to\min$，子序列 $x^{(t)}$ 收敛于一个局部最优解．

如果 Brown 虚拟对策法从含有非零元的混合策略 $x^{(0)}$ 开始，就是一种内点法．当 Karmarkar[K1]提出了一种对线性规划有很好收敛性的内点算法时，他便在数学规划领域完成了一个突破．此后，数以千计的出版物中提出了许多的改进和扩展，其中大部分处理的是凸规划．

即使对非空集合 S，S 中的内点集合 int(S)也可能是空集．S 的*边界*定义为满足如下条件的 $a\in R^n$ 的集合：对任何 $\varepsilon>0$，球 $|x-a|\leqslant\varepsilon$ 都包含 S 中的一点和 S 外的一点．注意当目标函数是仿射函数但不是常数时，每一个最优解（如果存在的话）属于可行域的边界．

一些内点集合 int(S) 为空集的数学规划可以变换为带有非空内点的规划并

由内点法求解. 例如，当 S 由约束 $g_i(x)\leqslant 0$，$i=1$，…，k($g_i(x)$是连续函数)给定时，我们可以用外点法得到一个点 y 使得对任何 i 和一个小 $\varepsilon>0$，有 $g_i(x)<\varepsilon$. 然后我们将约束 $g_i(x)\leqslant 0$ 放松到 $g_i(x)\leqslant\varepsilon$，使 S 扩大到 S_ε 而将这一点包含到 $\mathrm{int}(S_\varepsilon)$中去.

另一种方法对线性规划也有效. 通过求解某些线性方程组，并且除去可行域为 S 的线性规划中的一些变量，我们可以得到等价的可行域为 S' 的(由仿射变换得到)线性规划，使得或者两个规划都不可行，或者两个问题恰有一个可行解，或者 $\mathrm{int}(S')$非空. 更加复杂的方法(无需求解任何方程组)与矩阵对策有关. 任何线性规划的求解可化简为对称矩阵对策的求解，反过来矩阵对策也可化简为求解带有已知最优值(即 0)和内部有已知可行解的线性规划. [280]

许多无约束最优化或单变量最优化方法可以转换为内点法. 例如，对于坐标下降法已经清楚. 在梯度法或牛顿法中我们减小步长(如有必要)，可以使得到的解保持在 $\mathrm{int}(S)$中.

即使 $\mathrm{int}(S)$非空，找到一个可用的内点可能也不容易，这就是可行性问题或称为第 1 阶段问题. 内点法的修正方法(称为不可行内点法)，被建议用来处理这一问题.

现在我们考虑一些在[K1]之后变得尤其重要的方法.

仿射尺度法

令可行域 S 是凸的且由(A3.4)给定，$\mathrm{int}(S)$由 $g_i(x)<0$，$i=1$，…，m 给定. 我们从一点 $x^{(0)}\in\mathrm{int}(S)$开始. 给定任意一点 $w=x^{(t)}\in\mathrm{int}(S)$，给定 S 的约束可重写如下：

$$(g_i(w)-g_i(w))/g_i(w)\leqslant 1 \text{ 对 } i=1,\cdots,m.$$

含有 x_t 的区域 $S_t\subset S$ 定义如下：

$$\sum_{i=1}^{m}((g_i(w)-g_i(x))/g_i(w))^2\leqslant 1. \tag{A5.1}$$

其次定义 $F(w)=x^{(t+1)}$ 为 $f(x)$在 S_t 上的极小点. 在 $F(w)$碰到边界这一不太可能发生的情况下，我们可以做一些阻尼(即以 $w+\alpha(F(w)-w)$(正数 $\alpha<1$)替换 $F(w)$).

如果 $f(x)$在 S_t 上的极小化问题并不比在 S 上极小化简单时，这种方法实际上是没有用的. 一般来说，S_t 上的极小化问题并不一定更简单. 例如，当 $m=1$ 时，$S_t=S$ 对任意 t 都成立. 当$m\geqslant 2$ 时，集合 S_t 不一定是凸的. 然而，有一种重要的情形，此时 S_t 上的极小化问题更简单. 也就是，如果函数 $f(x)$和 $g_i(x)$是仿射的且 $f(x)=cx+d$ 不是常数(即 $c\neq 0$)，那么在定义 S_t 的约束(A5.1)中，左边的 $g(x)$是一个次数不大于 2 的多项式. 因此，我们可以很容易地找到 $x^{(t+1)}$. 一种方法是对变量做一个仿射变换，并且将 $g(x)$变为如下两个标准形式之一：$g(x)=z_1{}^2+\cdots+z_m{}^2+d_0$ 或者 $g(x)=z_1{}^2+\cdots+z_k{}^2+d_1z_{k+1}+d_0$，

$k \leqslant m-1$.

在第一种情形下，S_t 是新坐标系下的球 $|z|^2 \leqslant 1-d_0$（原坐标系下的椭球），而且在 S_t 上最小化 $f(x)=cx+d=\tilde{c}z+\tilde{d}$ 很容易，唯一的最优解是 $z^{(t+1)}=$ [281] $-(1-d_0 1)^{1/2}\tilde{c}/|c|$.

在第二种情形下，由于规划是有界的，而且对 $i>k$ 有 $\tilde{c}_i=0$，因此最优解 $z^{(t+1)}$ 的前 k 个元素是唯一的而且被赋予了一个相似的计算公式，第 k 个元素是任意的（当 $d_i=0$ 时）或者服从于一个线性约束，而其他元素（如果存在的话）是任意的.

我们不用改变变量，只需求解线性方程组（KKT 条件，见附录 A1）即可.

注意 $f(x)$ 仿射这一条件比较容易满足（见附录 A1），而且凸集 S 可由一个线性约束系统近似. 因此仿射尺度法可用于更加一般的凸规划，至少在原理上是这样. 例如，我们可以用线性约束 $g_i(x^{(t)})+g_i'(x^{(t)})(x-x^{(t)})\leqslant 0$ 来替代约束 $g_i(x)\leqslant 0$，可以利用阻尼使得到的解保持在 $\text{int}(S)$ 中.

实际的计算表明这一方法对初始内点 x_0 的选择很敏感. 一个好的技巧是保持远离边界. 有证据[H2]表明，如果我们从一个很接近最优解的点开始，或更一般地从很接近中心路径(central path)的点开始（见下面的讨论），这种方法对凸规划有很好的结果.

障碍法

令 S 是 R^n 的子集，且其内部 $\text{int}(S)$ 非空. S 的障碍函数 $B(x)$ 是 $\text{int}(S)$ 上的连续函数，对任何 $x\in\text{int}(S)$ 有 $B(x)\geqslant 0$，而且对任何收敛于 $\text{int}(S)$ 外部一点的任意序列 $z^{(t)}\in\text{int}(S)$ 有 $B(z^{(t)})\to\infty$. 注意除非 $S=R^n$，那么一个障碍函数不能扩展为 R^n 上的连续函数.

如果 S 由(A3.4)给定，下面是一些障碍函数 $B(x)$：

$$-\sum_i \log(-g_i(x))\,;\ -\sum_i -1/g_i(x))\,;\ -1/\max_i[g_i(x)].$$

保持 S 不动，$g_i(x)$ 的变化能生成更多的范例并消除这三个范例之间的差别.

给定带有连续函数 $f(x)$ 的有界数学规划(A1.1)，在非空集 $\text{int}(S)$ 上的障碍函数 $B(x)\geqslant 0$，以及满足 $\delta_t\to 0$ 的序列 $\delta_t>0$，我们用如下一系列规划来近似规划：

$$f(x)+\delta_t B(x)\to\min, x\in\text{int}(S), \tag{A5.2}$$

如果在 $\text{int}(S)$ 中开始，使用带有强下降性质的方法解决(A5.2)并忽略约束 $x\in$ [282] $\text{int}(S)$，那么我们会使得到的解保持在 $\text{int}(S)$ 中.

定理 A5.3 假设 $v=\inf\limits_{x\in\text{int}(S)}(f(x))>-\infty$，令 $v_t=\inf\limits_{x\in\text{int}(S)}(f(x)+\delta_t B(x))$，那么 $t\to\infty$ 时 $v_t\to v$.

证明 显然 $v_k\geqslant v_{k+1}\geqslant v$. 对任何 $\varepsilon>0$，我们找到 $y\in\text{int}(S)$ 使得 $f(y)-v\leqslant\varepsilon/2$. 其次我们找到 k 使得对 $t\geqslant k$ 有 $\delta_t B(y)\leqslant\varepsilon/2$，那么

$$|v_k - v| = v_k - v \leqslant (f(y) + \delta_t B(y)) - v \leqslant \varepsilon/2 + \varepsilon/2 = \varepsilon$$

对任意 $t \geqslant k$ 成立. ■

现在假设 $f(x)=cx$ 是线性的，$B(x)$是凸的，规划

$$f(x) + \delta B(x) \to \min, x \in \mathrm{int}(S)$$

有唯一最优解 $x^*(\delta)$，点 $x^*(\delta)$形成所谓的中心路径，因此障碍法也以中间道路法为人所知. 定理 A5.3 中的序列 $x^{(t)} = x^*(\delta_t)$遵循这一道路，从而有术语路径跟踪法(path-following method).

通常修正牛顿法从点 $w=x^{(t)}$开始的一步或多步可以找到点 $x^{(t+1)}$. 因此首先对 $f(x)+\delta_t B(x)$找到牛顿方向 $u=-B''(w)^{-1}(c/\delta_t+B'(w))^{\mathrm{T}}$，然后令 $x^{(t+1)}=w+\alpha u$，其中恰当选择数 α 以便改进 $f_k(x)$. $f(x)$的强下降性质可以自动保持 $x^{(t+1)}$在内部. 根据附录 A2，当 κ 是 $-(cu/\delta_t+b'(0))\,|b'''(0)|/b''(0)^2$ 和 $b(s)=B(w+us)$的上界时 $\alpha=1/(1+\kappa)$是一个好的选择.

更大的步长 α 可以使得到的解位于 S 之外. 不用 $f'''(x)$找到 α 的另一种选择是令$\alpha=\beta_t\alpha_t$，其中 α_t 是单变量规划$\alpha\to\max$，$x^{(t)}+\alpha z\in S$ 的最大值而且β_t 是阻尼序列(例如 $\beta_t=1/t$).

为了更好地收敛，我们希望 $B(x)$是自协调的且 $f_d''(x)\geqslant K_0>0$，那么 $f_k(x)$也是自协调的. 凸规划中的自协调障碍的存在在[NN]中得证. 另外，对规划中的一些类别，这样的障碍可以明确地构造出来(如同(A1.2)一样将目标函数变为线性之后).

对于由线性约束 $g_i(x)\leqslant 0$ 的有限系统给定的 S，一种好的自协调障碍是对数障碍 $B(x)=-\sum_i \log(-g_i(x))$.

障碍法的目的之一是当解接近边界时引入罚值，从而当使用无约束最优化方法时使得到的解保持在可行域中.

283

对此，通常要求在 int(B)中任意收敛于边界一点的序列 $z^{(k)}$有 $B(z^{(k)})\to\infty$. 然而，这会阻碍我们接近在边界的最优解，因此用到系数 $\delta_t\to 0$. 在某些情况下我们能找到并使用障碍函数，使得 int(S)中任何收敛于边界上非最优点的序列 $z^{(k)}$有 $B(z^{(k)})\to\infty$，而且 δ_t 是固定值.

路径跟踪法有一种形式也可称为滑动目标法或割平面法. 给定带有线性函数 $f(x)=cx$ 的凸规划(A1.1)，我们应用 S 的凸障碍函数 $B(x)$和正数序列 $\alpha_t\to 0$.

作为初始点，我们取 $x^{(0)}$为 $B(x)$在 S 上的极小值. 给定 $x^{(t)}$，对目标函数 $B(x)-\log(f(x^{(t)})-f(x)+\alpha_t)$(它是集合$\{x\in S,\ f(x)\leqslant f(x^{(t)})+\alpha_t\}$的凸障碍函数)应用一步(或者几步)修正牛顿法我们找到下一个点 $x^{(t+1)}$.

[NN]表明，对任何满足 int(S)非空有界的凸集 S 都有凸函数 $B(x)$，在(A4.7)的意义下它是关于 κ_2(依赖于 n)的自协调函数同时也满足(A4.9)意义下的 k_1. 当 $f(x)$仿射时，$f(x)+cp(x)$是关于同一个 κ_2 而满足不同的 k_1 的自协调函数.

对某个 S，在[NN]、[H2]中自协调障碍有明确构造. 在[NN]中允许自协调函数有不可逆的黑塞矩阵，需要在定义和牛顿法中作一些改变. 然而，在除去对 $f(x)$没有影响的变量以后，我们就回到了可逆黑塞矩阵的情形.

例 A5.4　S 由线性约束给定

$$g_i(x) \leqslant 0, i = 1, \cdots, m.$$

假设 $\text{int}(S)$非空，那么

$$B(x) = \sum_{i=1}^{m} \log(-g_i(x))$$

是凸障碍函数.

当 $\text{int}(S)$包含整条直线时，可以从规划中除去一个变量. 否则，对任给的 x，$f''(x)$都是可逆的. 当 S 有界时，条件(A4.10)对某 $K_2>0$ 成立.

从例 A2.16 中的计算可知，$p(x)$是关于 $\kappa_2=1$ 的自协调函数而且条件(A4.9)对 $\kappa_1=m^{1/2}$成立. ■

[284]

势函数化简法

给定数学规划(A1.1)，势函数(potential function)是 $\text{int}(S)$上满足如下条件的函数：对 $\text{int}(S)$中的任何序列 $x^{(k)}$，$f(x^{(k)})\to-\infty$当且仅当 $f(x^{(k)})\to f_0$，其中 f_0 是最优值. 因此 $f(x)$的最小化等价于 $h(x)$或 $e^{h(x)}$的最小化.

例如，令 $B(x)$为障碍函数，如果 x 收敛于一个最优解时有 $B(x)+C_0\log(f(x)-f_0)\to-\infty$，那么 $C_0\log(f(x)-f_0)+B(x)$是势函数.

现在我们考虑(A3.4)，其中 $f(x)$，$g_i(x)$为仿射函数. 在[K1]中，函数 $e^{h(x)}=(f(x)-f_0)^m/\prod_{i=1}^{m} g_i(x)$用于衡量下降的进程，但它不能直接决定移动的方向. 这是优函数的传统概念. 然而最小化势函数或优函数而不是最小化目标函数，一点都没错(除了术语上可能的混淆外).

但是，只有在某些最优化方法对 $h(x)$或 $e^{h(x)}$比对 $f(x)$更有效时以 $h(x)$或 $e^{h(x)}$替换 $f(x)$才有意义. 在[K1]中，$h(x)$不是凸的，但 $e^{h(x)}$是凸的而且自协调.

线性互补问题(LCP)

该问题是找到两个向量 x，$y\in R^n$ 使得

$$y = Ax + b, x \geqslant 0, y \geqslant 0, x_i y_i = 0(\text{对所有 } i),$$

A 是给定的 $n\times n$ 矩阵，向量 $b\in R^n$.

一个 LCP 是单调的(monotone)，如果矩阵 A 是半正定的. 由 KKT 条件，任何二次规划(特别地任何线性规划)可以以单调 LCP 计算.

人们提出了用于求解 LCP 的内点法(参见[FMP]).

半定规划(SDP)

考虑所有 $n\times n$ 实矩阵的集合 $M_n(R)$. 给定两个矩阵 $A=[a_{ij}]$，$B=[b_{ij}]$，

我们以 $A \cdot B$ 表示数 $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}$. 半定规划是 $C \cdot X \to \min$ 形式的数学规划，受约束于

$$A_{(i)} \cdot X = b_i (i = 1,\cdots,m), X \text{ 半正定},$$

其中 X 是对称 $n \times n$ 变量矩阵，对称矩阵 C 由数据构成，$A_{(i)} \in M_n(R)$ 和数 b_i.

SDP 在凸规划、组合最优化和控制论中有广泛的应用. 在[W1]，[K4]，[R]中，内点法用于求解 SDP.

285

A6. 扰动法

扰动法(pertubation)是第一种而且现在仍然是避免循环的最简单方法，它能保证单纯形法永远有效. 由于需要额外的计算，所以它并不实用. 但是，扰动法的理念在数学规划的其他领域很有用(见附录 A7). 在单纯形法的第 2 阶段(见第 10 节)，我们用列向量 $b(\varepsilon)=[b_1+\varepsilon,\ \cdots,\ b_n+\varepsilon^n]^{\mathrm{T}}$ 替代下表中的列向量 $b=[b_1,\ \cdots,b_n]^{\mathrm{T}}$：

$$\begin{matrix} x & 1 & \\ \begin{bmatrix} A & b \\ c & d \end{bmatrix} & & \begin{matrix} = u \\ = z \to \min, \quad x \geqslant 0, \quad u \geqslant 0 \end{matrix} \end{matrix}$$

形式上，我们处理关于 ε 的多项式. 非正式地，$b(\varepsilon)$是 b 的很小的扰动，ε 被看作一个小正数. 比较关于 ε 的多项式如下：

$$a_0 + a_1\varepsilon + a_2\varepsilon^2 + \cdots \geqslant f_0 + f_1\varepsilon + f_2\varepsilon^2 + \cdots, \text{如果}$$

$$a_0 > f_0, \text{或}$$

$$a_0 = f_0, \quad a_1 > f_1, \text{或}$$

$$a_0 = f_0, a_1 = f_1, \quad , a_2 > f_2, \text{或}$$

$$\cdots$$

现在，我们应用单纯形法，在第 2 阶段得到表

$$\begin{matrix} x & 1 & \\ \begin{bmatrix} A & b(\varepsilon) \\ c & d(\varepsilon) \end{bmatrix} & & \begin{matrix} = u \\ = z \to \min, \quad x \geqslant 0, \quad u \geqslant 0 \end{matrix} \end{matrix}$$

其中在初始表中 $d(\varepsilon)=d$，我们得到同样形式的表，而且 $b(\varepsilon)$的元素都不是 0. 这是因为 $b(\varepsilon)$的元素对实数总是线性无关的(它们在初始表中线性无关，并且接下来旋转步只引起这些元素的相加和实系数的乘法，因此它们保持线性无关). 所以元素 $d(\varepsilon)$的值经过每一旋转步都要减小，即 $d(\varepsilon)$不可能不变，这就使得循环是不可能的了. 如果得到最优的 ε 表，那么令 $\varepsilon=0$，我们就得到边值问题的最优表. 如果得到有一个“坏”列的 ε 表，那么令 $\varepsilon=0$，我们就得到边值问题有一个“坏”列的表.

286

A7. 目标规划

在实际生活中，可能有许多目标需要最优化．一个目标函数的最优解也是另一个目标函数最优解的情况并不经常发生．为了应用数学规划，我们需要将这些目标合并到一个目标函数中而且可能考虑对一些目标附加约束．有许多方法可以实现它，所以有一些关于目标规划、向量优化或者多目标规划的书(参见[CVL]，[K5]，[S])．类似的问题也出现在对策论中，其中不同局中人的目标是不同的．

假设要极小化函数 $f_1(x)$，…，$f_m(x)$．我们可以为每个目标设定数值目标 b_i 然后极小化函数的凸组合：

$$f(x)=\sum_{i=1}^{m}c_if_i(x)\to\min,f_i(x)\leqslant b_i\text{ 对所有 }i,x\in S,$$

其中对任何 i 有 $c_i>0$.

上述规划的每个最优解 x^* 是帕累托最优的(Pareto optimal)或者有效的(efficient)(即 x^* 是可行的，并没有其他的点 y 使得对任意 i 有 $f_i(y)\leqslant f_i(x^*)$且对某个 i 有 $f_i(y)<f_i(x^*)$)．这种点的集合成为有效前沿或者帕累托边界．在某些情形下，特别当 S 凸时，每一个帕累托最优解可作为 $f_i(x)$的凸线性组合的最优解而得到.

在 $k=n=2$ 的情况下，S 是凸的，$f_1(x)=x_1$，$f_2(x)=x_2$，纳什建议使用

$$f(x)=-(b_1-f_1(x))(b_2-f_2(x))$$

而不用

$$f(x)=c_1f_1(x)+c_2f_2(x).$$

他在二人合作对策中这么做，这时 x_i 是局中人 i 的支付.

S 的凸性确定了最优解的唯一性．如果有可行解且 S 为有界闭集，那么最优解 x^* 存在，这在对策论的背景下是自然而然的．此外，如果 $f(x^*)\neq 0$，那么 x^* 是帕累托最优的.

纳什方法的一个自然推广是如下规划：

287

$$f(x)=-\sum_{i=1}^{m}\log(b_i-f_i(x))\to\min,f_i(x)\leqslant b_i\text{ 对所有 }i,x\in S.$$

为得到新变量 $y_i=f_i(x)$下的凸规划，我们希望映射

$$x\mapsto[f_1(x),\cdots,f_m(x)]\in R^m$$

下 S 的像 S'是凸的．在对策论的背景下这可以通过采用混合策略得到．这等于用 S'的凸包替代 S'，因此问题的形式为

$$g(y)=-\sum_{i=1}^{m}\log(b_i-y_i)\to\min,y_i\leqslant b_i\text{ 对所有 }i,y\in S'.$$

其中目标函数是凸的和自协调的，这使得带有合适障碍函数的修正牛顿法有效地进行.

这些方法在各个目标重要性基本一致时使用. 在优先目标规划(preemptive goal programming)的情况下，各个目标函数根据优先级进行了部分排序. 如前所述，我们可以将处于同样优先级的目标合并. 但是，对不同优先级目标的处理是不同的.

例如，对 $f_i(x)$按优先级分类，并且假设 $f_1(x)$最重要. 我们可以利用如下的连续步骤求解全部问题，即求解 k 个数学规划. 首先我们最小化 $f_1(x)$得到最优值 $c_1{}^*$，然后我们最小化带有附加约束 $f_1(x)=c_1{}^*$ 的 $f_2(x)$(这只有在第一个规划有多于一个最优解时才有意义). 以此类推. 描述全部问题的一个好的方式是

$$f(x)=\sum_{i=1}^{k}\varepsilon^{i-1}f_i(x)\to\min, f_i(x)\leqslant b_i \text{ 对所有 } i, x\in S,$$

其中目标函数是关于 ε 的多项式而且对多项式如同在附录 A6 中一样比较. 注意现在目标函数不是实值的. 我们并不考虑多项式，而是考虑按字典顺序排列的 k 维行向量.

同时注意优先目标规划的任何最优解也是帕累托最优的.

合并 m 个目标函数 $f_i(x)$的一般方法是利用 m 变量的函数 $h(y)$，它在区域 $y\geqslant 0$ 中对每个变量都是非降的(对于可微情形，这意味着它在区域中有 $g'(y)\geqslant 0$). 那么我们的规划是

$$-h(b_1-f_1(x),\cdots,b_m-f_m(x))\to\min, f_i(x)\leqslant b_i \text{ 对所有 } i, x\in S.$$

这个规划的每一个最优解都是帕累托最优的. 在上述例子中，$h(y)=cy(c\geqslant 0)$而且$g(y_1, y_2)=y_1y_2$. 288

A8. 低维线性规划

对于较大规模的问题，最新的内点法胜过了单纯形法. 另一方面，对于带有较少变量(或者根据对偶性，较少约束)的线性规划，有比单纯形法更快的方法，参见[C2]. 我们将对两个变量 x，y 的线性规划演示这一过程.

以规范形式写这个规划

$$cx+c'y\to\min, ax+a'y\leqslant b, x\geqslant 0, y\geqslant 0,$$

其中 c，c' 是给定的数而a，a'，b 是给定的m 维列向量. 这个规划可以写成标准的行表

$$\begin{array}{c} \begin{matrix} \quad x & \quad y & \ 1 \end{matrix} \\ \begin{bmatrix} -a & -a' & b \\ c & c' & 0 \end{bmatrix} \begin{matrix} =*\geqslant 0 \\ \to\min \end{matrix} \end{array} \tag{A8.1}$$

可行域 S 至多有 $m+2$ 个顶点，从任何顶点开始的第 2 阶段至多在 $m+1$ 旋

转步里终止.

每一步旋转至多进行两次比较，为了检验表是否最优或者是否找到了旋转列向量. 然后，用 m 个符号检验，至多 m 个除法，至多 $m-1$ 次比较，我们找到一个转轴元或者一个“坏”列.

其次旋转，计算表中的 $3m+3$ 个新元素. 用一步除法，我们找到旋转行向量中的新元素(不包括最后一个元素，因为最后一个元素在此之前已经算出)而且用 $2m$ 次乘法和 $2m$ 次加法我们找到旋转行向量和列向量以外的新元素. 最后，在 $m+1$ 次除法里找到旋转列向量中的新元素. 因此一个旋转步包括找旋转元和旋转，可以在 $8m+3$ 次操作(operation)里完成——算术操作和比较操作.

因此，第 2 阶段可在 $(m+1)(8m+3)$ 次操作里完成. 尽管这一数字可以有所减少(例如在第 $(m+1)$ 个旋转步我们只需计算表的最后一列)，但任何单纯形法的修正都不可能使这个数字有实质的减少(在最坏的情况下).

289

就旋转步的数目来说，对任何 $m\geqslant 1$ 令 S 为有界凸 $(m+2)$ 边形显然可以，其中对任何顶点都有一个线性目标函数使得单纯形法恰需要 $\lfloor 1+n/2\rfloor$ 个旋转步，而且在每一步中旋转元只有一个选择. 也可以构造一个 $(m+2)$ 边形，一个目标顶点，以及一个初始顶点使得需要 m 个旋转步(每一步只有唯一的选择)(或者第一步有两个选择，第一个选择导致最优解而第二个选择导致 m 个额外的旋转步(每一步只有唯一的选择)).

现在我们概述寻找最优解的方法. 假设(A8.1)可行，则需要不超过 $100m+100$ 次操作. 下面对 m 归纳. 对于 $m\leqslant 12$ 可以使用单纯形法而且在

$$(m+1)(8m+3)<100m+100$$

次操作里完成. 因此我们假设 $m\geqslant 13$.

假设目标函数是非常数的(否则，我们只需两次比较即可). 令 $u=-cx-c'y$，当 $c'\neq 0$ 时 $v=x$，否则 $v=y$.

我们以如下形式重写所有的 $m+1$ 个条件(不包括 $v\geqslant 0$)：

$$u\leqslant a_i v+b_i, i=1,\cdots,l,$$
$$u\geqslant a_{-i} v+b_{-i}, i=1,\cdots,m+1-l.$$

这至多需要 $3(m+1)$ 次操作. 如果 $l=0$，那么规划无界，而且我们在小于等于 $3(m+1)\leqslant 100m+100$ 次操作里完成. 因此假设 $l\geqslant 1$.

如果数 $a_{2i-1}-a_{2i}$ 和 $b_{2i-1}-b_{2i}$ 有不同的符号或者对某个满足 $1\leqslant i\leqslant l/2$ 或 $1\leqslant -i\leqslant (m+1-l)/2$ 的 i 两个都是零，那么我们可以舍掉这两个约束之一. 这涉及到至少 $m+1$ 次减法和 $0.5m+0.5$ 次符号比较. 我们以 $l'\leqslant l$ 和 $m'-l'-1$ 分别表示余下的 $\leqslant$ 和 $\geqslant$ 形式约束的数目. 注意到 $l'\geqslant 1$. 如果 $m'\leqslant 12$，那么我们在

$$\begin{aligned}&3(m+1)+(m+1)+(0.5m+0.5)+100\cdot 12+100\\&=4.5m+4.5+100\cdot 12+12<100m+100\end{aligned}$$

次操作里完成. 因此假设 $m'\geqslant 13$.

我们将余下的约束写为

$$u \leqslant a'_i v + b'_i, i = 1, \cdots, l',$$

$$u \geqslant a'_{-i} v + b'_{-i}, i = 1, \cdots, m' + 1 - l'.$$

[290]

而且记住计算的数目 $q_i = a_{2i-1} - a_{2i}$，$p_i = b'_{2i-1} - b'_{2i}$，其中 $\mathrm{sign}(q_i) = \mathrm{sign}(p_i) \neq 0$.

现在计算 $1 \leqslant i \neq l'/2$ 和 $1 \leqslant -i \neq (m'+1-l')/2$ 对应的 $v_i = p_i/q_i$. 这给出

$$k = \lfloor l'/2 \rfloor + \lfloor (m+1-l')/2 \rfloor \leqslant (m'+1)/2$$

个数. 算术操作的数目是

$$k \leqslant (m'+1)/2 \leqslant 0.5m + 0.5.$$

然后计算这些 $k \geqslant 0.5\,m' - 1.5$ 个数的中位数 v_0，至多需要 $18k+18 \leqslant 9m+27$ 次比较.

其次找 $a_i' v_0 + b'(1 \leqslant -i \leqslant m'+1-l')$ 的最大值 u''，这需要 $m' - l'$ 次比较.

我们在计算 $\mathrm{sign}(a_i')$（满足 $a_i' x_i + b_i' = u'$）的集合 Y 的同时找 $a_i' v_0 + b'(1 \leqslant l')$ 的最小值 u'，这需要至多 $2l'-1$ 次比较. 操作的总数至多是

$$(4.5m + 4.5) + (0.5m + 0.5) + (9m + 27) + (m' - l') + (2l' - 1)$$
$$\leqslant 16m + 31.$$

如果 $u'' \leqslant u'$ 而且 Y 或者包含 0 或者包含 1 和 -1，那么 v_0 是最优解而且 u' 是最优值. 因此我们在 $(16m+31)+1 < 100m+100$ 次操作里完成.

如果 $u'' > u'$ 或者 $Y = \{-1\}$，那么对任何最优解 v^* 有 $v^* < v_0$. 当 $u'' \leqslant u'$ 和 $Y = \{1\}$，有最优解 $v^* \leqslant v_0$. 在两种情形下，我们都知道在 x_0 的哪一侧来搜索 v^*.

现在我们可以舍弃每一个与 v^* 不在 v_0 同侧的 v_i（包括 v_0）的约束，这意味着至少舍弃 $k/2$ 个约束，而至多

$$m' - k/2 \leqslant m' - (m' - 3)/4 = 0.75m' + 0.75 \leqslant 0.75m + 0.75$$

个保持不变. 由归纳假设，我们可以在至多 $100(0.75m+1.75)$ 次操作里完成. 因此，在至多

$$(16m + 31) + 1 + 100(0.75m + 1.75) = 91m + 207 < 100m + 100$$

次操作里，可以解决我们的线性规划.

如果进行更加精细的考虑，数字 $100m+100$ 可以得到改进.

[291]

A9. 整数规划

本节讨论带有附加条件的线性规划，即某些变量是整数. 一个特别的情况是布尔变量，即二进制变量，或在组合规划中，要求所有的变量是 0 或 1. 一般地，可以将变量写为以 2 为底（二进制表示）的方式将所有变量都是有界整数情形化简为二进制情形. 在组合规划的情况下，一个解可被认为是一个变量的集合（取值 1）. 如果所有的变量都要求是整数，我们就有纯整数规划.

例 A9.1 maximize $f(n)$ subject to $1 \leqslant n \leqslant N$，$n$ 是整数，其中 N 是给定的

正整数.

这是一个单变量整数规划. 我们希望找到有限序列 $f(1)$, $f(2)$, …, $f(N)$ 中的最大项. 该问题可由 $N-1$ 次比较解决(见下面附录 A10). ■

例 A9.2 (工作指派问题) 这是一个 0-1 规划, 可以化简为线性规划. ■

例 A9.3 (背包问题)

$$cx \to \max, ax \leqslant b, x = [x_1, \cdots, x_n]^{\mathrm{T}},$$

其中整数 $x_i \geqslant 0$, a, $c \geqslant 0$ 是给定的行向量而 b 是给定的数.

这个整数规划考虑的是重量为 b 的背包所能装下的最大价值, 其中 x_i 是装下价值为 c_i 重量为 a_i 情形下物品 i 的数目. 在有界背包问题中, 我们有额外的约束 $x_i e_i$. 在 0-1 规划中, 所有的 $e_i = 1$. ■

例 A9.4 (旅行商问题) (参见[GP].)

给定一个 $n \times n$ 费用矩阵 $[c(i, j)]$, 一个环游(tour)就是城市(city)1, 2, …, n 的一个置换 $[\sigma(1), \cdots, \sigma(n)]$. 一个环游意味着恰访问每个城市一次, 然后返回第一个城市(称为家). 环游的费用是总费用

$$c(\sigma(1), \sigma(2)) + c(\sigma(2), \sigma(3)) + \cdots + c(\sigma(n-1), \sigma(n)) + c(\sigma(n), \sigma(1)).$$

旅行商问题是找总费用最小的一个环游. 它可表述为(如同工作指派问题中那样)二进制变量 x_{ij} 的整数规划. ■

292

利用二进制变量, 我们可以将逻辑上比较复杂的约束系统化简为所有约束都要求满足的通常的系统. 例如, 规划

$$\text{minimize } f(x) \text{ 满足三条件之二}: g_1(x) \leqslant 0, g_2(x) \leqslant 0, g_3(x) \leqslant 0$$

可写为

$$f(x) \to \min, y_1 g_1(x) \leqslant 0, y_2 g_2(x) \leqslant 0, y_3 g_3(x) \leqslant 0,$$

$$y_1 + y_2 + y_3 = 2, y_i \text{ 为二进制变量}.$$

下面我们讨论求解整数规划的一些方法.

舍入线性解

去掉变量为整数的条件所得到的线性规划, 称为整数规划(IP)的线性松弛规划(LP-relaxation). 如果线性规划的一个最优解满足整数限制(例如工作指派问题), 那么它对 IP 就是最优的. 一般地, 对线性规划的最优解进行舍入, 得到 IP 的一个"解". 这个解不必可行, 而且即使它可行, 也不必最优. 在现实生活中如果更好的解很难找到有时就用这个"解".

穷举和随机枚举

当限制为整数的变量有界时, 通过跑遍这些变量所有可能的值将 IP 化简为线性规划的有限集合. 这种方法只有当线性规划的变量数目较小时才实用. 一般地, 我们随机地选择值. 尝试的选择越多, 找到 IP 的最优值的概率就越接近 1.

使用凸包将 IP 归结为线性规划

考虑 IP 的可行域 S 以及它的凸包 S'(即 S 内所有点的所有凸组合的集合).

S'可由线性约束的有限集合给定，而且目标函数 $f(x)$在 S 上的最优化等价于 $f(x)$在 S'上的最优化. 此外，S'的每一个最优的极点都属于 S 从而对 IP 也是最优的. 尽管理论上这个方法可将任何 IP 归结为一个线性规划，但是只有当 S'由少量的线性约束描述时这一方法才有效. 见第 3 节的例子.

分支定界法

我们使用最好的分支优先规则概述整数规划(带有 0-1 变量 x_1，…，x_n). 实际上它可用于某些变量是整数的任何数学规划，但有不同的分支规则. 我们从 IP 的最小化规划 IP_0 的线性松弛规划 LP_0 开始(即将 IP_0 中的条件 $x_i^2=x_i$ 替换为条件 $0\leqslant x_i\leqslant 1$ 而得到的线性规划). 如果 LP_0 的最优解是整数，问题结束. 在任何情况下 LP_0 的最优值都是 IP_0 的最优值的一个下界. 下面我们将 IP_0 分裂为两个 IP，即固定 $x_1=0$ 或 1 得到 IP_1 和 IP_2. 我们求解相应的线性规划得到 IP_1 和 IP_2 的最优值的下界 v_1 和 v_2. 然后我们将带有下界的规划分支成两个规划，对应$x_2=0$ 和$x_2=1$. 继续这一过程，我们得到一棵树在第 t 步有 t 个节点. 在每个节点通过求解相应线性松弛规划我们得到相应 IP 的下界. 如果带最小下界的节点其最优解都不是整数，我们就将其中一个节点分支成两个. 这一过程至多在 2^n-1 步分支内终止.

[293]

割平面法

我们从求解 IP 问题的线性松弛规划开始，如果最优解满足整数约束，问题结束. 否则，增加额外的线性约束，割去松弛规划的最优解，但整数规划的所有可行解(或者至少所有的最优解)均满足这个约束. 重复这一过程，得到可行域逐步缩小的线性规划序列. 构造割平面约束的许多方法已被提出，可以证明这一过程在有限步内终止. 然而，步数很大的情况也很常见. 多面体合并(polyhedral annexation)是一种寻找已知处于多面体一个极点 P 的最优解的割平面法. 一般的算法是从某个极点出发求解多面体合并问题. 结果是或者确认极点是(全局)最优的，或者可以生成使用凸割除去极点的后续方向. 这种方法通过割去区域生成一个逐步收缩的多面体序列. 它的名字来自于从极点解出发，排除原始多面体的某些区域，形成一个新多面体的思想.

线性规划的很多教科书包含整数规划的内容，而且也有专门介绍整数规划的书(参见[ES]，[W2]，[KV]，[S2]).

[294]

A10. 排序和顺序统计量

人类和计算机都进行大量的排序. 排序是数据处理的重要部分，因此找到有效的排序方法非常重要. 这里排序意味着在一个线性表中对项目排序，比如一个班级中的学生列表. 更确切说，给定 $n\geqslant 1$ 个数 a_1，…，a_n，则 $b_1\leqslant\cdots\leqslant b_n$ 是同样数的排序.

我们希望尽快找到这些数 b_i. 通常，任何排序方法都可找到一个置换 σ 使得

$b_{\sigma(i)}=a_i$.

排序涉及比较数字和可能来回移动这些数字. 我们只对比较计数，为了简便我们称之为“步”. 因此我们的问题就是如何以最小的步数对一个给定的数字集合排序.

因此，我们只对收集有关数字相对大小的信息成本感兴趣，而忽略存储和使用这些信息的成本以及算法的复杂性.

既然 n 个数字有 $n!$ 种置换，众所周知任何排序方法至少需要$\lceil\log_2 n!\rceil$步，其中$\lceil x\rceil$代表最小的整数 $t\geqslant x$. 我们证明如下广为人知的结果(参见[K3]).

定理 A10.1 可以用

$$F_0(n)=\sum_{i=1}^{n}\lceil\log_2 i\rceil=n\lceil\log_2 n\rceil-2^{\lceil\log_2 n\rceil}+1$$

步完成对 n 个数的排序.

通过对 n 归纳，只需证明下面的结论即可.

引理 A10.2 给定一列排好序的数 $b_1\leqslant\cdots\leqslant b_{n-1}$ 和另一个数 a_n，至多需要$\lceil\log_2 n\rceil$步完成对所有 n 个数的排序.

证明 换言之，我们必须证明只需 k 步足以将一个新的数 b_n 插入 $n=2^k-1$ 个数的一个有序列表里. $n=k=1$ 时显然只需一步. 令 $k\geqslant 2$，使用与附录 A2 中相似的二分法继续对 k 做归纳. 也就是，将 a_n 与中位数 $b_{(n+1)/2}$ 比较，然后将 a_n 插入已经排好序的$(n-1)/2=2^{k-1}-1$ 个数的有序列表里，由归纳假设这可在 $k-1$ 步里完成. ■

[295]

显然，$k-1$ 步是找到 $\min=\min[a_1,\cdots,a_n]$的充分必要条件，对于最大数 max 这也成立.

如果我们还有关于序列 a_i 的额外信息，就可以用它来更快地找到 min. 例如，假设序列是单峰的(unimodal)(即有一个 i^* 使得 a_i 在 $i\leqslant i^*$ 严格递减而在 $i\geqslant i^*$ 严格递增). 一种用最小计算步数找 $\min=a_{i^*}$ 的方法，类似于斐波那契搜索法，称为格点搜索(lattice search). 它比在$\lceil\log_2 n\rceil$步里找到最小值的二分法更有效.

下面仍然假设我们对 a_i 一无所知. 虽然需要大约 $\log_2 n!$ 步才能找到 b_i 的有序列表，但是对任意特殊的 b_i(顺序统计量)，包括中位数 $b_{\lceil n/2\rceil}$，可以更快地算出，也就是用 Cn 步就已足够，其中 C 是不超过 n 和 k 的常数，见[K3].

对较小的 C，现有的证明都很长，因此我们给出几个例子然后概述 $C=18$ 时的一个简单证明.

例如，$b_i=\min(a_i)$可由 $n-1$ 步比较计算得到.

另一个例子，考虑特殊情形 $n=5$，排序至少要用$\lceil\log_2(5!)\rceil=8$ 步. 另一方面，正如前面所述，要用 $n-1=4$ 步(比较)找到 b_1. 然后用三步找到 b_2(余下的数中的最小值). 因此用七步可找到 b_2. 类似地，用四或七步可找到 b_5 和 b_4. 最

后，用如下七步可找到中位数 b_3.

首先用插入法，用 $1+2+2=5$ 步将 a_1，a_2，a_3，a_4 排序并得到 $c_1\leqslant c_2\leqslant c_3\leqslant c_4$. 然后我们用 a_5 与中位数 c_2 和 c_3 比较(再用两步). 如果 $c_2\leqslant a_5\leqslant c_3$，那么 $b_3=a_5$. 如果 $a_5\leqslant c_2$，那么$b_3=a_2$. 如果 $a_5\geqslant c_3$，那么 $b_3=a_3$. 因此，每个 b_i 都可在七步内找到.

从[DZ1]中知道，任何统计量 b_i 都可以在至多 $2.95n+o(n)$步内找到，而在最坏的情形[DZ2]下，至少需要 $2.01n+o(n)$步. 这里 $o(n)$(或 $o(1)$)表示满足 $n\to\infty$时 $o(n)/n\to 0$[或$o(1)\to 0$]的序列.

现在我们简述只需 $18(n-1)$步就足以找到第 k 小的数 b_k. 继续对 n 归纳. $n\leqslant 34$ 的情形显然，因为我们可以在 $18(n-1)$步内对 n 个数排序，因此令 $n\geqslant 35$. [296]

设 $n=10l+5+r$，$0\leqslant r\leqslant 9$. 将前 $10l+5$ 个数分解为 $2l+1$ 个 5 元组并且找到每一个 5 元组中的中位数 c_1，…，c_{2l+1}，这可以在 $7(2l+1)$步内完成. 由归纳假设，我们可以在 $36l$ 步内找到 c_j 的中位数 d.

现在我们有 $3l+2$ 个数 a_i 在 d 的右边，有 $3l+2$ 个数 a_i 在 d 的左边. 在 $n-6l-5=4l+5$ 步内我们将余下的 $n-6l-5$ 个数 a_i 放在 d 的右边或左边.

现在我们必须在 d 右边的 $n'\leqslant n-3l-3$ 个数中，或者在 d 左边的 $n''\leqslant n-3l-3$ 个数中找到一个确定的统计量. 由归纳假设，我们可以在 $18(n-3l-4)$步内完成.

因此，总的步数至多是

$$7(2l+1)+36l+4l+5+18(n-3l-4)=18n-60\leqslant 18n-18.$$

最后我们讨论在一个给定 $m\times n$ 矩阵$[a_{ij}]$中找鞍点的问题(或者证明它们不存在). 当我们试图求解矩阵对策时会出现这种问题，而且它可以描述成整数规划.

给定 i，j 后，需要用 $m+n-2$ 步检验这个位置是否为鞍点. 另一方面，需要用 $n(m-1)$步寻找每一个列向量中的所有极大元. 类似地，要用 $m(n-1)$步寻找每一个行向量中的所有极大元. 因此 $2mn-m-n$ 步后问题结束(被两次选中的位置就是鞍点). 人们还不知道是否有更快的方法.

但是，有更快的找到严格鞍点的方法(即对任意 $k\neq j$ 和 $l\neq i$ 的(i,j)有 $a_{ik}>a_{ij}>a_{il}$)(或者证明严格鞍点不存在). 方法[BV]用 $F_0(m)$步对主对角线上的数 a_{ii}开始排序(假设 $m\leqslant n$)，而在

$$F_0(m)+F_0(m+1)+n+m-3+(n-m)\lceil\log_2(m+1)\rceil$$

步内终止. [297]

A11. 其他问题及最新进展

我们还没有提到数学规划中的一些重要问题，包括一些特殊的数学规划的类别以及求解这些规划的特殊方法.

除了明确给出的数以外，数据可能还包括参数或外部变量(比如线性函数的系数)或随机数据. 在14节中，我们提到了线性规划背景下的参数规划. 随机规划(Stochastic programming)(参见[BL]，[KW])专门处理数据的不确定性(这也是模糊规划(fuzzy programming)所主要考虑的)(参见[C1]，[RI]，[RV]). 模糊集也用于表示与目标规划相联系的优先权.

为求解大规模的问题，需要特别复杂的技巧处理数据，而且要修正这些方法以便在合理的时间内得到一个好的解. 现在有一些最新的关于大规模最优化的书：[B2]，[T]. 对于大规模的线性规划问题，有单纯形法的一些修正形式，其中数据处理吸引了特别的注意. 例如，在表中列向量数目特别大的情况下，可以在旋转的过程中需要它们时再生成. 列生成和割平面法是对偶的. 而且，单纯形法也有特殊的修正形式处理变量的上界.

一些大规模的线性规划可以分解为彼此弱相关的子规划. 这会引出如同Dantzig-Wolfe方法一样的分解法(嵌套规划). 类似地，整体规划法可将求解一个大的线性规划问题化简为求解一系列小的线性规划问题.

在分数规划(fractional programming)中，目标函数 $f(x)$ 和约束函数是 $a(x)/b(x)$ 形式的和，其中 $a(x)$，$b(x)$ 是仿射函数且在 S 上有 $b(x)>0$. 在只有一项的情况下，问题看起来像

$$a_0(x)/b_0(x) \to \min, a_i(x)/b_i(x) \leqslant d_i \text{ 对 } i=1,\cdots,m,$$

其中 $a_i(x)$，$b_i(x)$ 都是仿射函数，d_i 是常数. 这个情况比较特殊，因为这类规划等价于一个线性规划(假设在可行域中所有的 $b_i(x)\leqslant 0$). 见[C-M].

在可分规划(separable programming)中，目标函数 $f(x)$ 和每一个约束函数都是单变量函数 $f_i(x_i)$ 的和. 分段近似和线性规划用于求解这类规划(参见 [298] [S3]).

在多项式规划中，(A1.9)中的目标函数 $f(x)$ 和约束函数 $g_i(x)$ 是多项式. 线性规划对应于次数为1的情况. 处理多项式函数比较容易(例如计算它们的导数)，而且某些问题，例如可行解或最优解的存在性，理论上可以在对有理数据的有限多次算术运算内得到回答(但是在高次多变量的情况下我们不知道如何有效地做到这一点). 然而，求解多项式规划不会比求解带有连续函数的规划问题容易多少，因为有界区域上的连续函数可由多项式近似.

运输问题(见第6章)是各种各样网络优化问题中的一个特殊情况，其中的一些问题可以归结为运输问题，而且一些问题又具有整数数据一定导致整数最优解这一性质. 线性规划的很多教科书都论述网络问题，而且有关于网络问题的专门书籍，包括非线性网络问题(参见[ES]). 内点法如今也用于求解大型的网络问题.

动态规划(参见[DL])是考虑一段时间内的最优决策. 对于连续时间，它用在最优控制和变分学中. 对于离散时间，我们有多阶段(多期)模型. 多期决策的

主要思想是从最后阶段开始求解问题，按时间回溯. 位置对策(也叫对策的推广形式)也应用这一方法(参见[FSS]，[M]).

线性规划的一个很流行的应用是数据包络分析(data envelopment analysis)，见 http：//www. banxia. com/，http：//www. deazone. com/.

神经网络(Neural networks)通过模仿生物神经系统，可用于求解一些最优化问题("学习"算法)，而且数学规划可用于设计有效的神经网络.

迭代法可能依赖于一些参数. 为一个带有难于计算或者我们知之甚少的目标函数或者带有一个复杂可行域的数学规划，选择参数是个令人畏缩的使命. 一种方法是遗传(genetic)或进化(evolutionary)规划. 我们从一些算法出发，这些算法称为编码串，带有随机选择的参数和初始点. 对每个编码串进行一些迭代，根据相应目标函数的改进对编码串排序. 坏编码串被去掉，好编码串则被配对，而 299 它们的"后代"出现时，一些参数值进行交换(交叉)，或者随机地改变一点点(突变)，或者以某些方式组合. 如果与并行计算(多 CPU)结合，这个方法就特别地有吸引力. (见[CVL]，[LP].)

一般地，硬件的进步(比如并行计算机，量子计算机和 DNA 计算机)也会刺激产生数学规划中的新方法(参见[CP2]，[DPW]，[G]，[NC]，[LP]，[MCC]). 300

参考文献

[B1] Berkovitz, L. D., *Convexity and Optimization in R^n*. J. Wiley, New York, 2002.

[B2] Biegler, L. T., et al., *Large-Scale Optimization with Applications.* Springer-Verlag, New York, 1997.

[BL] Birge, J. R. and F. Louveaux, *Introduction to Stochastic Programming.* Springer Series in Operations Research. Springer-Verlag, New York, 1997.

[BV] Byrne C. and L. N. Vaserstein, An improved algorithm for finding saddlepoints of two-person zero-sum games, *Game Theory,* 20:2 (1991), 149–159.

[C1] Carlsson, C.,*Fuzzy reasoning in decision making and optimization.* Physica-Verlag, Heidelberg/New York, 2002.

[C2] Chazelle, B., *The Discrepancy Method. Randomness and Complexity.* Cambridge University Press, Cambridge, U.K., 2000.

[CGT] Conn, A. R., N. I. M. Gould, and P. L. Toint, *Trust-Region methods.* SIAM, Philadelphia, 2000.

[CP1] Coope, I. D. and C. J. Price, On the convergence of grid-based methods for unconstrained optimization. *SIAM J. Optim.* 11 (2000), no. 4, 859–869

[CP2] Calude, C. S. and G. Paun, *Computing with Cells and Atoms: An Introduction to Quantum, DNA, and Membrane Computing.* Taylor & Francis, London/New York, 2001.

[CVL] Coello, C. C. A., D. A. Van Veldhuizen, and G. B. Lamont, *Evolutionary algorithms for solving multi-objective problems.* Kluwer Academic, New York, 2002.

[D] Dorn, W. S., Duality in quadratic programming, *Quarterly of Applied Mathematics* 18 (1960), no. 2, 155–162.

[DL] Dreyfus, S. E. and A. M. Law, *The Art and Theory of Dynamic Programming.* Mathematics in Science and Engineering, Vol. 130. Academic Press [Harcourt Brace Jovanovich, Publishers], New York/London, 1977.

[DPW] Du, D.-Z., P. M. Pardalos, and W. Wu (ed.), Mathematical Theory of Optimization. Kluwer Academic, Dordrecht/Boston, 2001.

[DZ1] Dor, D. and U. Zwick, Selecting the median, *SIAM J. Comput.* 28 (1999), no. 5, 1722–1758.

[DZ2] Dor, D. and U. Zwick, Median selection requires $(2+\epsilon)N$ comparisons. *SIAM J. Discrete Math.* 14 (2001), no. 3, 312–325.

[ES] Eiselt, H. A. and C.-L. Sandblom, *Integer Programming and Network Models.* With contributions by K. Spielberg, E. Richards, B. T. Smith, G. Laporte and B. T. Boffey. Springer-Verlag, Berlin, 2000.

[FMP] Ferris, M.C., O. L. Mangasarian, and J.-S. Pang (editors), *Complementarity: Applications, Algorithms, and Extensions.* Kluwer Academic, Boston, 2001.

[FV] Florenzano, M. and C. L. Van in cooperation with P. Gourdel, *Finite Dimensional Convexity and Optimization* . Springer-Verlag, Berlin/New York, 2001.

[FSS] Forg, F., J. Szp, and F. Szidarovszky, *Introduction to the Theory of Games. Concepts, Methods, Applications.* Revised and expanded version of the 1985 original. Nonconvex Optimization and its Applications, 32. Kluwer Academic, Dordrecht/Boston, 1999.

[FW] Frank, M. and P. Wolfe, An algorithm for quadratic programming. *Naval Res. Logist. Quart.* 3 (1956), 95–110.

[G] Gramss, T. et al., Non-Standard Computation: Molecular Computation, Cellular Automata, Evolutionary Algorithms, Quantum Computers. Wiley-VCH, Weinheim/New York, 1998.

[GP] Gutin G. and A. P. Punnen (editors), *The Traveling Salesman Problem and its Variations* Kluwer Academic, Dordrecht/Boston, 2002.

[H1] Hačijan, L. G., A polynomial algorithm in linear programming. (Russian) *Dokl. Akad. Nauk SSSR* 244 (1979), no. 5, 1093–1096.

[H2] Hertog, D. den., *Interior Point Approach to Linear, Quadratic, and Convex Programming: Algorithms and Complexity.* Kluwer Academic, Dordrecht/Boston, 1994.

[HT] Horst, R. and H. Tuy, On the convergence of global methods in multiextremal optimization. *J. Optim. Theory Appl.* 54 (1987), no. 2, 253–271.

[K1] Karmarkar, N., A new polynomial-time algorithm for linear programming. *Combinatorica* 4 (1984), no. 4, 373–395.

[K2] Kelley, C. T., *Iterative Methods for Optimization* (Frontiers in Applied Mathematics, 18) SIAM, Philadelphia, 1999.

[K3] Knuth, D. E., *The art of computer programming. Volume 3. Sorting and searching.* Addison-Wesley Series in Computer Science and Information Processing. Addison-Wesley Publishing Co., Reading, Mass., 1973.

[K4] Klerk, E. de, *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications.* Kluwer Academic, Dordrecht/Boston, 2002.

[K5] Kalyanmoy D., *Multi-Objective Optimization Using Evolutionary Algorithms.* John Wiley & Sons, Chichester/ New York, 2001.

[KV] Korte, B. H. and J. Vygen, *Combinatorial Optimization: Theory and Algorithms,* 2nd ed. Springer-Verlag, New York, 2002.

[KW] Kall, P. and S. W. Wallace, *Stochastic programming.* Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Ltd., Chichester, 1994.

[L] Luenberger, D.G., *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, Reading, Mass., 1984.

[LP] Langdon, W. B. and R. Poli, *Foundations of Genetic Programming.* Springer-Verlag, New York, 2002.

[M] Morris, P., *Introduction to Game Theory.* Springer-Verlag, New York, 1994.

[MCC] Marathe, A., A. E. Condon, and R. M. Corn, On combinatorial DNA word design. *DNA Based Computers, V* (Cambridge, Mass, 1999), 75–89, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 54, Amer. Math. Soc., Providence, RI, 2000.

[NC] Nielsen, M. A. and I. L. Chuang, *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, U.K., 2000.

[NN] Nesterov, Yu. and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, Philadelphia, 1994.

[P] Porembski, M., Finitely convergent cutting planes for concave minimization, *Journal of Global Optimization,* 20, no. 2 (June 2000), 113–136.

[R] Renegar, J., A Mathematical View of Interior-Point Methods in Convex Optimization. SIAM/Mathematical Programming Society, Philadelphia, 2001.

[RI] Ramík, J. and M. Inuiguchi, *Fuzzy Mathematical Programming.* Papers from the session of the 7th Congress of the International Fuzzy Systems Association held in Prague, June 25–29, 1997. Edited by Jaroslav Ramk and Masahiro Inuiguchi. Fuzzy Sets and Systems 111 (2000), no. 1. North-Holland Publishing Co., Amsterdam, 2000.

[RV] Ramík, J. and M. Vlach, *Generalized Concavity in Fuzzy Optimization and Decision Analysis.* Kluwer Academic, Boston, 2002.

[S1] Schniederjans, M. J., *Goal Programming: Methodology and Applications.* Kluwer Academic, Dordrecht/Boston, 1995.

[S2] Sierksma, G., *Linear and Integer Programming: Theory and Practice.* 2nd ed. Marcel Dekker, New York, 2002.

[S3] Stefanov S.M., *Separable Programming: Theory and Methods.* Kluwer Academic, Dordrecht/Boston, 2001.

[S-M] Stancu-Minasian, I. M., *Fractional Programming: Theory, Methods and Applications.* Kluwer Academic, Dordrecht/Boston, 1997.

[SMY] Sarker R., M. Mohammadian, and X. Yao (editors), *Evolutionary Optimization* . Kluwer Academic Publishers, Boston, 2002.

[T] Tsurkov, V., *Large-Scale Optimization: Problems and Methods.* Applied Optimization, 51. Kluwer Academic, Dordrecht/Boston, 2001.

[VCS] Vaserstein, L. N., V.I. Chmil', and E. B. Sherman, A multiextremal problem for the growth and location of production with a concave objective function. (Russian), in *Mathematical Methods in Economic Research* (Russian), pp. 138–143. Izdat. Nauka, Moscow, 1974.

[V] Vaserstein, L. V., On the best choice of a damping sequence in iterative optimization methods, *Publ. Matem. Univ. Aut. Barcelona* 32 (1988), 275–287.

[W1] Wolkowicz, H. et al (editors), *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications.* International Series in Operations Research & Management Science, 27. Kluwer Academic, Dordrecht/Boston, 2000.

[W2] Wolsey, L. A. *Integer Programming.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, 1998.

索　引

索引中的页码为英文原书页码，与书中页边标注的页码一致.

E

F

G

H

I

J

K

L

M

N

O